ELECTRICAL COMMUNICATION

Arthur Lemuel Albert, M.S., E.E.
Professor of Communication Engineering
Oregon State College

THIRD EDITION

John Wiley & Sons, Inc., New York
Chapman & Hall, Limited, London

THIRD EDITION

Second Printing, May, 1954

PRINTED IN THE UNITED STATES OF AMERICA

PREFACE TO THIRD EDITION

As was true of the preceding editions, the third edition covers the entire field of electrical communication, including the transmission of code, speech, and music by both wire and radio.

The third edition has been extensively revised. The basic subjects of acoustics, electroacoustic devices, networks, lines, cables, wave guides, and electronics have been grouped in the first part of the book. Telegraph, telephone, and radio systems have been placed in the last part. Both radio systems and dial telephone systems have received much greater emphasis than in previous editions.

Covering the entire field of modern electrical communication in a single book has necessitated a careful selection of material. It is regretted that more text material and more illustrations could not be included. However, the extensive lists of references will assist those interested in locating additional information.

An important feature of the third edition is the addition of review questions at chapter endings; also, the problems have been revised and increased in number. The standards of the electrical and radio professions have been followed closely.

The many courtesies of the American Telephone and Telegraph Company, Pacific Telephone and Telegraph Company, Bell Telephone Laboratories, and Western Electric Company have been greatly appreciated. Also courtesies extended by International Telephone and Telegraph Company, Western Union Company, Radio Corporation of America, Brush Development Company, General Radio Company, Racon Electric Company, and others are gratefully acknowledged.

Appreciation is expressed to the editors of *Electrical Engineering, Proceedings of the Institute of Radio Engineers, Bell System Technical Journal, Bell Laboratories Record, Electrical Communication, Electronics, Tele-Tech, Telephony,* and others, for permissions to use quotations, illustrations, and other material. Appreciation is also expressed to the authors of the many technical articles from which information was obtained.

The assistance of Mr. Dwight L. Jones, of the Pacific Telephone and Telegraph Company, who prepared most of the material on dial telephone systems, is acknowledged with gratitude.

ARTHUR L. ALBERT

OREGON STATE COLLEGE
April, 1950

PREFACE TO FIRST EDITION

Although several excellent books treating specialized phases of electrical communication are available, there is a need for a book which considers these subjects as they are related to one another in modern communication systems. This book, accordingly, presents the various phases of transmitting intelligence electrically as they contribute to the end desired—that of providing the public with an adequate and economical communication service.

This book is designed primarily for use as a college textbook, and as a reference book for those in the communication industry who have had a college education or other technical training. However, those not having such training will find much of value included.

Although higher mathematics has been employed where advisable, the reader not prepared to follow these solutions will find, with few exceptions, that the text is easy to follow. In most chapters the use of mathematics has been limited; care has been taken, however, to insure that this has not been accomplished at a sacrifice of engineering exactness.

In preparing the manuscript the most recent communication standards have been rigidly followed. The material is thoroughly modern in every detail.

Engineering educators will recognize that a high percentage of the electrical engineering graduates now in the communication industry are engaged in engineering work which is broad and non-technical in nature. Relatively few graduates are engaged in such theoretical work as filter design or transmission studies. In selecting material for the manuscript, this fact has been considered. It has been the aim of this book, instead of merely presenting the electrical theories of communication, to include a discussion of the entire industry and thus provide a basic training upon which a successful career in communication engineering can more readily be built.

Those in the communication industry who are engaged in administrative, supervisory, or similar activities which are essentially non-technical in nature, but who desire a better understanding of the plant and engineering features of their industry, should find this book of great value.

The manuscript has been reviewed by many unbiased authorities on communication, and every care has been taken to insure its correctness. It is with the greatest appreciation that this assistance is acknowledged.

The writings of many authorities have been consulted during the preparation of the manuscript. Many of these articles are listed at the end of the chapters, and these have been freely referred to wherever it is felt that the reader would be benefited. These references are especially valuable to the reader who is not well acquainted with the subject. Many quotations are included, particularly where it seemed desirable that the original wording should be maintained. The contributions of these many writers are most gratefully acknowledged.

The courtesies of the Bell Telephone Laboratories, the American Telephone and Telegraph Company, the Radio Corporation of America, the Raytheon Production Corporation, and the American Institute of Electrical Engineers, in supplying descriptive material and in rendering other assistance, is greatly appreciated.

The assistance of the students and faculty of the school of engineering at Oregon State College has been invaluable. In particular, the interest and cooperation of H. S. Rogers, President of Brooklyn Polytechnic Institute and former Dean of Engineering at Oregon State, and of R. H. Dearborn, Professor and Head of the Department of Electrical Engineering and Acting Dean of the School of Engineering at this institution, have made this work possible.

Arthur L. Albert

Oregon State College

CONTENTS

CHAPTER 1

EARLY HISTORY OF ELECTRICAL COMMUNICATION

Introduction. It is surprising how little was done to improve methods of communication from primitive times until about the middle of the nineteenth century. Before this time, with the exceptions of the signal fire, semaphore, and heliograph, it was necessary for a person to travel from place to place and communicate verbally, or for a written message to be carried between two points. Methods of instantaneous communication had to await the development of electrical knowledge, and this development was slow and laborious.

This chapter will be devoted to a study of the events which led to the perfection of electrical methods of communication. These historical events well illustrate the early problems and also provide a foundation upon which the later theoretical considerations can be built.

Early Views of Electricity and Communication. Before the Christian era, electrical knowledge was confined to the facts that rubbed amber attracted light objects and that lodestone or magnetite exerted magnetic effects. It is stated[1] that Homer noted these facts in the twelfth century B.C. In 1267, Bacon published his theories of the polar attraction of lodestone, which did much to stimulate thought along electrical lines.

After the appearance of Bacon's work, rumors were circulated about a "certain sympathetic needle" which could be used to transmit information over long distances. This needle is thought to have been first described[1] in print by Porta in 1558. According to the rumors, which were unfounded, if both needles were magnetized from the same lodestone by rubbing, a movement of one needle, even though at a great distance, was supposed to cause a similar sympathetic movement in the needle of the other instrument. In this manner, it was supposed, communication could be established.

[1] Superior numbers refer to the numbered references at the end of the chapter, page 15. The "1," for instance, refers the reader to the articles on the telegraph and telephone in the Encyclopædia Britannica.

Early Attempts at Electrical Communication. The earliest known attempts[2] at transmitting electricity over a distance were by Grey in 1727 and Dufay in 1733. The Leyden jar gave an impetus to such experiments, and in 1747 Watson succeeded in discharging a jar and measuring the effect over almost two miles of iron wire supported on poles. Franklin performed similar experiments in 1748, but neither he nor Watson attempted communication in this manner.

It is of interest to consider the extent of electrical knowledge at this time. Gilbert had found by 1600 that a large number of substances when rubbed attracted light objects. In 1629 Cabeo described his theory of electric repulsion in electrified bodies. The first electrical machine, consisting of a sulfur ball rubbed with the hand, was developed by von Guericke in 1650. This fundamental invention was further perfected until it was possible to generate very high voltages by frictional methods. The battery and the electrical generator employing electromagnetic induction, however, were not yet available.

The early methods of transmitting intelligence were by telegraph systems, the word "telegraph" being derived from the Greek words *tēle* meaning distant, and *graphein* meaning to write. Morrison in 1753 is credited with having suggested the first telegraph system. A separate circuit was to be used for each character to be signaled. A frictional generator producing a high voltage was to be connected at the sending end, and small pieces of paper would be attracted by suitable electrodes at the distant station. Some years later this method was put into operation, but like all the devices based on electrostatic attraction it was of little value.

The first attempt at electrical telegraphy was made by Lessage in 1774, using a system similar to that suggested by Morrison. Letters were signaled by the divergence of pithballs. A revolutionary idea was introduced by Lomond in 1787; he used a similar system but employed *one wire* and a *code* of signals. Other methods employing spark discharges at the receiving end to identify the signals were introduced by Reusser (1794) and Cavallo (1795).

All these methods are chiefly of historic interest; with the possible exception of the idea of a code as introduced by Lomond, these early systems contributed little to the present telegraph.

The Development of the Electromagnetic Telegraph. In 1800, Volta produced the battery and, for the first time, made continuous currents available for experimentation. A current, as distinguished from electrostatic methods, was probably first employed to transmit intelligence by Salva in 1805, and by Soemmering in 1809. Both these men determined the signal at the receiving end by passing the current

between two electrodes immersed in water, and by detecting the presence of gas bubbles.

The electromagnetic telegraph starts with Oersted, who in 1819 found that a magnetized needle would be deflected from its normal position when brought close to a wire carrying a current. A year later, Schweigger found that the deflection was increased if the needle was wrapped with several turns of wire. Soon after, Ampère studied electromagnetism and proposed the use of magnetic needles and coils for the reception of signals.

Ampère's first telegraph system employed a pair of line wires for each character to be sent, which made it of little value. Schilling in 1832 exhibited a system similar to Ampère's, but using fewer line wires. Gauss and Weber in 1833 made further improvements, employing a suspended bar magnet with an attached mirror which reflected a beam of light as a signal detector. The signal currents employed in their apparatus were generated by electromagnetic induction, independently discovered by both Faraday and Henry. Much credit is due Steinheil, who highly perfected the systems of Gauss and Weber, and who discovered that the earth could be used for one side of the connecting circuit. Steinheil also arranged for the messages to be received by several different means, including a code of dots recorded on paper moved by clockwork, acoustically by means of needles which struck bells, and visually by observing the motion of needles.

Wheatstone and Cooke did much to popularize the telegraph and to demonstrate its commercial practicability. Their first system was of the deflecting-needle type, requiring six line wires. It was used with railway systems but was superseded by a type having only two line wires. The independent discoveries of Argo and Davy that an electric current had power to magnetize steel and the development of the electromagnet by Sturgeon in 1825 contributed greatly to the progress of telegraphy. Henry improved the electromagnet and in 1821 successfully used this principle to operate a magnetized bar which rang a bell and transmitted intelligence by code. It remained, however, for Samuel F. B. Morse to apply this principle in a satisfactory manner.

The Invention of the Electromagnetic Telegraph. Morse was born in Charlestown, Massachusetts, in April, 1791, and was graduated from Yale University in 1810. He was interested in both science and art. In fact, it was while returning from a study of art in Europe that Morse conceived the principle of his telegraph. On board the ship in October, 1832, a fellow passenger showed Morse an electromagnet and performed a number of experiments with it. Also, Morse learned that the speed of electricity was considered almost instantaneous.

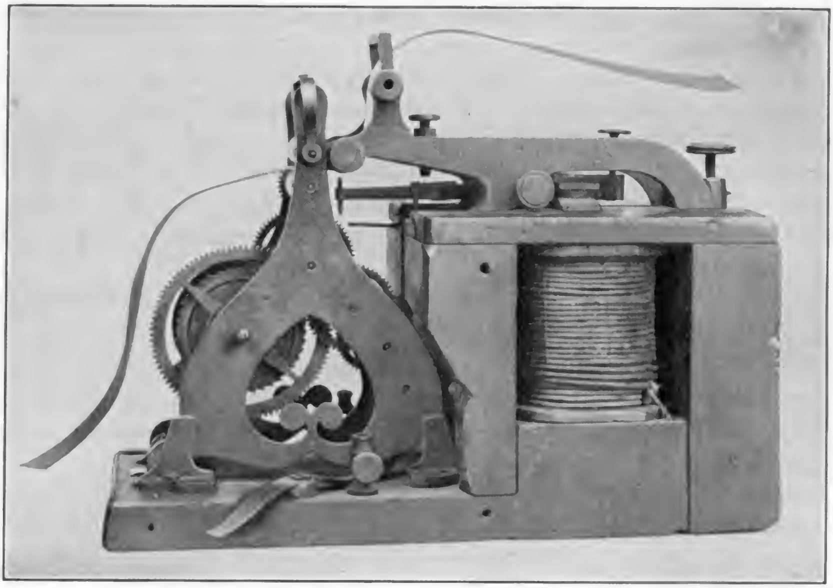


Fig. 1. An early Morse telegraph recorder.

He concluded that, if he could arrange a satisfactory detecting device, signals could be rapidly transmitted between distant points.

In an early instrument shown in Fig. 1, Morse had a stylus attached to a movable armature which was actuated by the pull of an electromagnet. The stylus traced a record of the dot and dash code impulses on a moving piece of paper. About 1835, Morse privately demonstrated his telegraph and obtained the backing of Gale and Vail. These men obtained financial aid and later made improvements. The early sending switch was operated by drawing beneath it either a notched bar or a bar in which pegs were arranged in accordance with a code.

After traversing a few miles of line, the received currents were too weak to operate the receiving apparatus. Morse, with the aid of Gale and Henry, devised the relay system shown in Fig. 2. The received currents passed through a coil, and an armature was actuated, closing a battery circuit and sending over the outgoing line an impulse from this local battery many times more intense than the signal received.

Further Development of the Morse Telegraph System.[3,4] Morse's early instrument operated by mechanical means without the assistance of a trained telegraph operator. Soon the system was

modified by Morse and his associates, and a telegraph key, to be operated by hand, was substituted for the sending apparatus previously described. The method of reading the Morse dot and dash code message from sound rather than from a tape was perfected by Vail in 1844.

Early in 1838, Morse applied for his patents, which were granted in 1848. In 1838 he demonstrated his apparatus before the President of the United States and his cabinet, and a sum of money was provided by the government for the construction of a line between Washington and

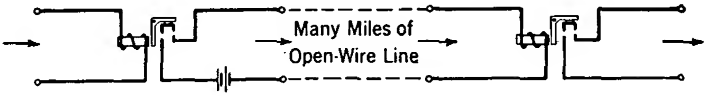


FIG. 2. An early telegraph relay system.

Baltimore. Over this circuit were sent, in 1844, the now-famous words "What hath God wrought!" When Morse offered his telegraph to the government it was refused on the recommendation of the postmaster-general,[1] as he was "uncertain that the revenues could be made to equal its expenditures."

It is not feasible to record here all the subsequent developments that followed Morse's invention. Among the more outstanding are the various methods of increasing the message-carrying capacity of the line. Gintl invented the **duplex** system in 1853, and Heaviside in 1873 and Edison in 1874 independently invented the **quadruplex** system. The **multiplex** system was suggested by Farmer in 1852, and later developed by Meyer (1873), Baudot (1881), and Delany (1884). Recent developments in telegraphy will be considered in Chapter 9.

The Development of Submarine Telegraphy.[5] The possibility of submarine telegraphy was suggested by Salva in 1798. It was not, however, until the introduction of gutta-percha as an insulator that submarine cables became of commercial importance.

One of the first attempts to lay a submarine cable was made in 1850, when a cable was placed between England and France. Because of its weak construction, this cable broke shortly after communication was established. In 1851 a second attempt was made and this cable proved successful.

The first attempt to lay a cable across the Atlantic was made by Field and his associates in 1857, but this ended in complete failure. In 1858 a cable was laid between Newfoundland and Ireland and was

operated for about three months, after which it failed. In 1865 an effort was made to lay another cable, which broke when about two-thirds completed. The following year, however, a cable was laid across the ocean, and this cable proved satisfactory. There are now more than twenty cables across the Atlantic between North America and Europe, and more than 3500 submarine telegraph cables in the entire world, with a total length exceeding 350,000 miles.

Early Attempts to Transmit Speech Mechanically. In the following pages the methods of transmitting intelligence by the voice rather than by codes will be considered.

The word "telephone," which is derived from the two Greek words *tēle* meaning distant, and *phōnē* meaning sound, is thought[6,7] to have been first used by Huth in 1796. This usage was in connection with a plan for relaying the spoken word over a distance by megaphones and ear trumpets. Early attempts at transmitting speech were mechanical rather than electrical; that is, the sound waves were directed by trumpets or through tubes, or the sound impulses were carried through rods or wires between the two communicating points. It is probable that speaking trumpets and ear trumpets date back to the distant past and that these were the first method of increasing the distance of speech transmission. Although references are made[6] to "monstrous trumpets of the ancient Chinese," these devices were the object of experimentation in comparatively recent years.[8] Speaking tubes, which are used today, were discussed[6] as early as 1589, although they probably date back much farther.

A second method of transmitting speech during this early period consisted of using rods, strings, or wires to conduct the sound impulses between the two points. As early as 1665, Hooke[6] discussed this subject in some detail. Early mechanical telephones were developed by connecting two distantly located sounding boards by means of some intervening object such as a rod, a wire, or a string.

The **string telegraph** or **lovers' telegraph** is the best example of speech transmission by mechanical means. The origin of this instrument is uncertain,[6] but it probably dates back several hundred years.

Early Attempts to Transmit Speech Electrically. Bourseul, in 1854, is considered the first to have attacked the problem of transmitting speech by electrical means. In fact, Bourseul nearly stated the correct principle of the telephone about twenty years before it was actually invented. The mistake he made was to suggest that the "disk alternately makes and breaks the current from a battery." Had Bourseul suggested a variable contact for the make-and-break contact, success might have been his.

One of the first inventors to attempt to transmit speech electrically was Reis,[9] who in 1861 constructed a device which he called a telephone. He apparently followed the suggestions of Bourseul, as he had a transmitter consisting of a stretched membrane operating a make-and-break contact. The action of the receiver was based on the Page effect; that is, if a rod of iron is suddenly magnetized or demagnetized a faint sound is given out. This device did succeed in transmitting sound of one pitch but did not transmit speech.

The Invention of the Electric Telephone. The telephone by which speech was first transmitted electrically was invented by Alexander Graham Bell. He was the descendant of a line of men famous in speech, elocution, and acoustics. His grandfather, Alexander Bell, was a professor of elocution.[10] His father, Alexander Melville Bell, followed this same calling in Edinburgh and London. He specialized in treating stammering pupils.

For some years young Bell was a teacher of elocution in various schools, and during this period he met in London two men who greatly influenced his life.[10,11] These men were Ellis and Wheatstone. Ellis told Bell of the work of Helmholtz with vibrating tuning forks driven by electromagnets, and it is probable that this first inspired Bell with the idea of a musical telegraph. Wheatstone demonstrated various devices such as the talking machine.

To seek a more healthful climate for young Bell, the family left England and settled in Canada. In 1871 he went to Boston, where he taught a class of deaf mutes. He was later made a professor in Boston University and opened his own School of Vocal Physiology.

Bell's early efforts were directed toward a harmonic or musical telegraph. As his investigation proceeded, he became convinced that the human voice could be conveyed by electrical means. Bell made a statement which showed that he had conceived the correct solution to the problem. According to Watson, who was Bell's assistant in the investigations, he said:[12] "If I could make a current of electricity vary in intensity, precisely as the air varies in density during the production of a sound, I should be able to transmit speech telegraphically." About this same time Bell visited Henry, who was then foremost in electrical science in America. He explained his ideas and demonstrated certain experiments to Henry, and was told[7] that he possessed "the germ of a great invention." Bell remarked that he felt that he did not have the electrical knowledge required to overcome certain difficulties and was advised by Henry to *"get it."*

It was on June 2, 1875, that the first sounds were transmitted by Bell and Watson as they were experimenting with their harmonic tele-

graph apparatus shown in Fig. 3. The manner in which it occurred can best be told in Watson's own words:[13]

I had charge of the transmitters, as usual, setting them squealing one after the other, while Bell was retuning the receiver springs one by one, pressing them against his ear as I have described. One of the transmitter springs I was attending to stopped vibrating and I plucked it to start it again. It didn't start and I kept on plucking it, when suddenly I heard a shout from Bell in

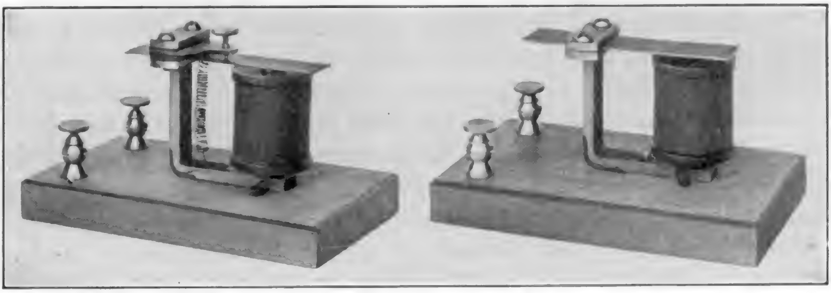


FIG. 3. Bell's vibrating reed transmitter (left) and receiver (right).

the next room, and then out he came with a rush, demanding, "What did you do then? Don't change anything. Let me see!" I showed him. It was very simple. The contact screw was screwed down so far that it made permanent contact with the spring, so that when I snapped the spring the circuit had remained unbroken while the strip of magnetized steel by its vibration over the pole of its magnet was generating that marvelous conception of Bell's—a current of electricity that varied in intensity precisely as the air was varying in density within hearing distance of that spring.*

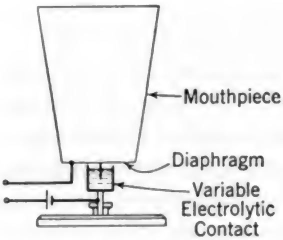


FIG. 4. Simplified diagram of Bell's liquid transmitter.

Bell immediately gave Watson directions for constructing the first telephone.

This instrument was not satisfactory, and when it was tried only a few faint tones were heard. Bell recognized that the generated currents from his transmitter were very weak, and accordingly he devised the liquid or electrolytic transmitter shown in Fig. 4. The current from a battery was caused to vary in accordance with the sound waves striking the transmitter by the variable liquid contact. It was with an instrument of this type that on March 10, 1876, Bell spoke the first sentence,[13] *"Mr. Watson, come here, I want you."*

When exhibited at the Centennial Exposition in Philadelphia, the

* Reprinted by permission, courtesy the American Telephone and Telegraph Company.

telephone attracted very little attention and was even ridiculed. It happened, however, that at the time the judges were about to leave without trying Bell's set, a fortunate incident occurred.[11] The Emperor of Brazil, Dom Pedro de Alcantara, walked in and greeted Bell. The Emperor had met Bell at Boston University, where he visited one of Bell's classes of deaf mutes. The Emperor examined the telephone receiver, and Bell went to the transmitter. Bell spoke into the transmitter, and to the utter amazement of the onlookers, the Emperor exclaimed: "My God—it talks." The telephone immediately became the center of attraction.

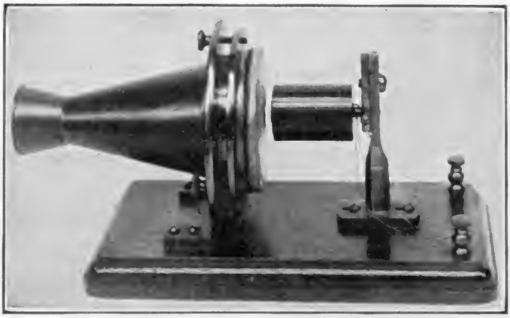


FIG. 5. Bell's original Centennial magneto transmitter.

The Early History of the Telephone. Early telephone instruments consisted[7] of a coil of wire on a permanent magnet near which was hinged a piece of iron. The iron moved in accordance with the vibrations of a tightly stretched parchment diaphragm. In later sets the hinged piece was not used; instead, the diaphragm had a soft-iron disk attached to its center. As the diaphragm and the attached iron vibrated under the influence of the sound waves, a feeble voltage was induced in the coil, and this voltage caused a current to flow in the distant receiver. Later, an iron diaphragm was substituted for the parchment. The sets exhibited in Philadelphia are shown in Figs. 5 and 6.

At first the public was reluctant to accept the telephone, and it was only after a number of demonstrations that orders were received. These demonstrations often were carried on over the then-existing telegraph lines, one of the longest transmissions being from New York to Boston. The first interconnection of telephone-equipped lines through a switchboard was made in Boston in 1877. These lines belonged to the Holmes Burglar Alarm Company.[14] The first commercial switchboard was placed in service in New Haven in 1878, serving eight lines and twenty-one subscribers.

As would be expected, there were a number of persons who contended that they, and not Bell, had invented the telephone. Such controversies were carried to the United States Supreme Court, where it was decided that Bell was the inventor of the telephone. The most formidable opponent of Bell and his associates was the Western Union, a company which had been operating telegraph lines for a number of years. This organization had both Edison and Gray on its staff. It is of interest to note that Gray had filed a notice that he had an *idea* upon which he intended to conduct further investigation with the United States Patent Office only a few hours later on the same day that Bell filed his patent application.

FIG. 6. Bell's original Centennial receiver.

In 1877, Berliner[15] applied for a patent on a transmitter. As Fig. 7 shows, this transmitter consisted essentially of a diaphragm which was free to vibrate under the influence of sound waves, and against which rested a metal ball. When the diaphragm vibrated, the resistance at the contact with the ball was varied but was not broken. In this manner the current from a battery was caused to vary in intensity as the sound waves varied.

Edison introduced his transmitter in 1878. This worked on the same principle as Berliner's, but employed contacts consisting of a platinum disk on the diaphragm and a fixed button of carbon.

A number of interesting experiments involving loose contacts in transmitters were performed[16] by Hughes about 1878. He demonstrated that loose-contact transmitters, or **microphones** as he called them, were superior to the firmly made contacts of the Edison and Berliner transmitters. It remained for Hunnings, however, to make the greatest contribution. He introduced a variable-resistance transmitter consisting (Fig. 8) of a mass of finely divided carbon granules instead of a solid carbon contact. The carbon granules give a large number of points of microphonic contact and are used today in almost all telephone transmitters. During this same period Berliner made an important contribution to telephony by proposing the use of an induc-

tion coil with the transmitter (see page 347). Edison was the first to use this principle commercially.

Blake offered to the Bell interests an improved microphonic transmitter consisting of a platinum and a carbon electrode held in contact with springs. This proved successful and was willingly accepted. The Western Union Company became convinced that the Bell patents were valid and withdrew from the telephone business.

The Development of Telephone Service. The first telephone switchboards were modeled after the then-existing telegraph boards. In some of these the party to be called was signaled by a buzzer ar-

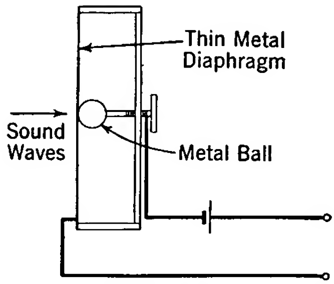


FIG. 7. Simplified diagram of Berliner's transmitter.

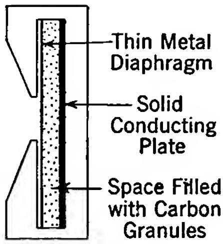


FIG. 8. Simplified diagram of the transmitter designed by Hunnings.

rangement causing the distant receivers to give out a grating sound.[17] This receiver, in addition to fulfilling the signaling function, also served as a magneto transmitter. The "Universal" and "Standard" switchboards were introduced in 1879 and 1880 respectively. Other switchboards and many important changes were adopted in the years following.[7,17]

Until about 1891 almost all telephones were operated with local batteries at each set to supply the transmitter current, and with hand generators for signaling purposes. An important advance was the common-battery switchboard developed[18] from the fundamental inventions of Carty. This switchboard employed one battery at the central office for all the telephone sets connected. The first common-battery switchboard was installed in Lexington, Massachusetts, in 1893. The so-called automatic systems (now called **dial telephone systems**) early received the attention of inventors, and in 1879 a patent on such apparatus was applied for by the Connellys and McTighe.[17] Many interesting switching systems were developed during this period.[19] The **step-by-step** dial system was invented by Strowger about 1887 and patented in 1889. It was developed by

Keith and others and was further perfected by Bell System engineers. This equipment was first used commercially at LaPorte, Indiana, in 1892. The **panel-type** dial system was developed by the Bell System early in this century. The **rotary** dial system was developed at about this same time and is widely used abroad. The **crossbar** dial system was developed about 1935.

The first circuits were burglar alarm, telegraph, or similar poorly adapted lines.[20] In most instances the earth was used as one conductor, and such lines are usually noisy (Chapter 14). Furthermore, the conductors employed were usually iron wire which offered high impedance to the voice currents. Hard-drawn copper wire first was used for telephone conductors in 1883 and has since *almost completely* supplanted iron in open-wire telephone lines.

The development about 1881 of the two-wire **metallic circuit** eliminating the earth return is credited to Carty. Carty did much of the early work on inductive interference[18] and developed methods for transposing circuits in 1891. He is also considered the inventor[11] of the **phantom circuit.** The first transposition system was devised by Barrett, in 1885.

The introduction of cable was an event of importance. It is stated[7] that Bell early conceived the idea of cables. Probably the first cables used for telephone service[7] were laid across Brooklyn Bridge in 1879. These were insulated with rubber or gutta-percha. Some early cables were insulated with oil. The insulation of all these cables had a high dielectric constant, and the cable capacitance was accordingly higher than desirable. Cotton cloth impregnated with paraffin was also used to insulate the wires. In the early cables the cores were first made and then pulled into the cable sheath. About 1890 the first "dry-core" cables were made, the insulation for each conductor consisting of a dry paper tape wound around the conductor. These insulated wires were formed into a core, and the lead sheath then pressed around the core by a continuous process. About 1940 a transcontinental telephone cable system was completed.

One of the greatest of all improvements in telephone transmission was the development and application of **loading** in 1899 and the years immediately following. The theory of loading is largely due to Heaviside, Pupin, and Campbell. Inductance was added to the lines at regular spacings to improve the transmission characteristics, and this application greatly extended the talking distances. Prior to the use of loading, commercial service could be given over no greater distance than from New York to Chicago; with loading, service was possible from New York to Denver by 1911.

Two important developments were the telephone **repeater** and the **carrier** telephone system. The vacuum-tube amplifier was invented by de Forest; it was adapted to telephone-line repeater operation by Bell System engineers, particularly by Arnold. The repeater using vacuum tubes was preceded by the Shreeve mechanical repeater, which, although it was used in a limited way, was not entirely satisfactory. It was the vacuum-tube repeater (first used experimentally[21] in 1913) that made possible the operation of the first transcontinental telephone circuits opened for commercial service in 1915. This was perhaps the first extensive commercial use of the vacuum-tube amplifier. The carrier-telephone system (sometimes called a multiplex system) was developed by Bell System engineers, and was first used in 1918.

The Development of Radio or Wireless Communication. From one standpoint, radio is not a different form of communication. As with other systems, radio involves sending intelligence either in the form of a telegraphic code or by the human voice as electromagnetic-wave variations. The early contributors were numerous, and it is difficult in a few paragraphs to trace the history adequately.[22,23,24,25,26]

Bell and Tainter in 1878 were probably the first to succeed in telephoning by "wireless" means.[10,24,25] Their device, called a "photophone," used a beam of light for conveying the talking impulses between two points.

Henry was among the first to experiment with electrical action at a distance without connecting wires.[27] By 1843 he had succeeded in magnetizing needles at a distance of 220 feet from the source.[26] In 1865 and 1873 Maxwell gave his famous papers on electromagnetic phenomena and showed mathematically that electric action is propagated through free space in the form of waves which travel with the velocity of light. In 1869 Ruhmkorff perfected a coil which for a number of years was used almost exclusively to supply the radio-frequency voltage for the antennas. In 1879 Hughes gave a demonstration of the transmission and reception of radio signals over distances of about 60 feet. He also discovered the principle on which the **coherer**—a detecting device later perfected by Branly—operated in wireless reception.[24,26] In 1883 Dolbear and in 1885 Edison both developed signal systems employing an elevated antenna and a ground connection. Also, in 1883 Edison applied for a patent on a two-electrode vacuum tube.

Hertz ranks with Maxwell in making outstanding contributions to the progress of radio communication. In 1881 Hertz published[1] an account of his experiments; he showed that rapid electrical oscillations

in conductors caused disturbances in space and that these had the characteristics of waves. Hertz worked with waves about 3 meters long and demonstrated that they could be reflected, directed, refracted, and polarized.[26]

The Establishment of Radio Communication. As previously mentioned, the **Ruhmkorff coil** was used for experimental work as a source of antenna excitation. In 1892 Thomson developed an arc method of producing high-frequency currents, and in 1896 Tesla applied for patents on a synchronous **rotary spark gap.** During this year Rutherford succeeded in receiving signals over a distance of half a mile, and Marconi successfully conducted tests over one and three-fourths miles.

The year 1897 was important in radio history: Marconi filed for patents on a wireless system, and Lodge developed the theory of tuning as applied to radio circuits. Marconi, perhaps more than any other, deserves the credit for placing radio communication on a commercial basis. During this same year signals were transmitted from shore to a ship 18 miles away. By 1900 radio telegraphy was becoming reasonably well established and was rapidly proving its great value. In 1901 Marconi received signals across the Atlantic.

The coherer, previously mentioned, was used as a detecting device for many years. In 1902 Marconi invented an improved **magnetic detector** which superseded the coherer and had wide application until the advent of the **two-electrode thermonic tube,** the **crystal,** and the **electrolytic detectors.** The two-electrode tube is of special importance and was employed as a detector by Fleming in 1904. In 1906 de Forest invented the **"audion"** or triode **vacuum-tube amplifier,** an event outstanding in the history of communication.[28] The Marconi interests initiated regular transatlantic telegraph service in 1907.

The systems thus far traced in this brief consideration were for telegraphic communication only. Radio telephony is based on the work of Fessenden, who in 1902 developed a system of modulating radio frequencies with the voice. In 1915 engineers of the Bell System succeeded in telephoning from Arlington to Hawaii, and from Arlington to Paris.[25] In 1920 the first commercial radio stations were installed for connecting two land telephone systems. This was between the mainland and Santa Catalina Island off the coast of California. In 1927 a transatlantic radio-telephone circuit was opened providing a connection between the telephone systems of Europe and America. Other radio links with important European centers followed. Similar telephone connections have been made from the United States to South

and Central America, to Bermuda, to Hawaii, and to the Orient. European countries have extensive commercial radio service with distant trade centers. Many ships at sea are connected by radio for commercial two-way telephone service. Similar service is provided with certain trains. Experimental conversations have been held with aircraft in flight to determine the possibilities of establishing regular commercial telephone service. Radio-telephone connections are provided in certain areas between coastal and harbor craft and between motor vehicles and the telephone network.

Radio broadcasting was inaugurated[29] by Conrad in 1920, using a system of amplitude-modulated waves. The so-called **superheterodyne** radio-receiving set, so widely used since about 1930, was developed by Armstrong. He also perfected a system of frequency modulation[30] in about 1936.

From the early dreams and investigations of Alexander Graham Bell and other communication pioneers has grown a vast international communication system, serving to bind together the scattered peoples of the world. This system may fulfil the vision of a distinguished telephone engineer, the late John J. Carty, who said:

"I have the faith that we shall some day build up a great world telephone system making necessary to all nations the use of a common language which will join all the peoples of the earth into one brotherhood."

REFERENCES

1. *Encyclopædia Britannica,* "Telegraph" and "Telephone."
2. *A Dictionary of Applied Physics,* Vol. 2, "Electricity." The Macmillan Co.
3. Bibber, H. W. The Centenary of the Morse Telegraph. *Elec. Eng.,* December, 1944, Vol. 63, No. 12.
4. d'Humy, F. E., and P. J. Howe. American Telegraphy After 100 Years. *Trans. A.I.E.E.,* 1944, Vol. 63.
5. Pratt, H., and J. K. Roosevelt. Developments in the Field of Cable and Radio Telegraph Communications. *Trans. A.I.E.E.,* 1944, Vol. 63.
6. Miller, K. B. *Telephone Theory and Practice (Theory and Elements).* McGraw-Hill Book Co.
7. Rhodes, F. L. *Beginnings of Telephony.* Harper & Bros.
8. Dyer, F. L., and T. C. Martin. *Edison: His Life and Inventions.* Harper & Bros.
9. Harlow, A. F. *Old Wires and New Waves.* D. Appleton-Century Co.
10. Mann, F. J. Alexander Graham Bell—Scientist. *Elec. Eng.,* March, 1947, Vol. 66, No. 3.
11. Casson, H. M. *The History of the Telephone.* A. C. McClurg & Co.
12. Watson, T. A. How Bell Invented the Telephone. *Trans. A.I.E.E.,* 1915, Vol. 34.

13. Watson, T. A. *The Birth and Babyhood of the Telephone.* Information Dept., American Telephone and Telegraph Co.
14. Thompson, G. K., and R. B. Hill. The First Telephone Switchboard and Its Method of Operation. *Bell Telephone Quarterly,* July, 1930, Vol. 9, No. 3.
15. Wile, F. W. *Emile Berliner—Maker of the Microphone.* Bobbs-Merrill Co.
16. Frederick, H. A. The Development of the Microphone. *Bell Telephone Quarterly,* July, 1931, Vol. 10, No. 3.
17. Jewett, F. B. The Telephone Switchboard—Fifty Years of History. *Bell Telephone Quarterly,* July, 1928, Vol. 7, No. 3.
18. John J. Carty—A Biographical Note. *Bell Lab. Record,* September, 1930, Vol. 9, No. 1.
19. Smith, P. C. Curious Patents in Mechanical Switching. *Bell Lab. Record,* March, 1929, Vol. 7, No. 7.
20. Gray, G. H. The Evolution of Wire Transmission. *Elec. Commun.,* 1942, Vol. 20, No. 4.
21. A Quarter-century of Transcontinental Telephone Service. *Bell Telephone Quarterly,* January, 1940, Vol. 19, No. 1.
22. Fleming, J. A. *The Principles of Electric Wave Telegraphy.* Longmans, Green & Co.
23. Erskine-Murray, J. *A Handbook of Wireless Telegraphy.* D. Van Nostrand Co.
24. Blake, G. G. *History of Radio Telegraphy and Telephony.* Chapman & Hall, Ltd.
25. Espenschied, L. The Origin and Development of Radiotelephony. *Proc. I.R.E.,* September, 1937, Vol. 25. Also *Bell Telephone Quarterly,* October, 1937, Vol. 16.
26. Morse, A. H. *Radio: Beam and Broadcast.* D. Van Nostrand Co.
27. Osborne, H. S. Joseph Henry—Pioneer in Electrical Science. *Elec. Eng.,* November, 1942, Vol. 61, No. 11.
28. de Forest, L. The Audion. *Elec. Eng.,* March, 1947, Vol. 66, No. 3.
29. Mourmtseff, I. E. A Quarter Century of Electronics. *Elec. Eng.,* Feb. 1947, Vol. 66, No. 2.
30. Armstrong, E. H. A Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation. *Proc. I.R.E.,* May, 1936, Vol. 24, No. 5.

CHAPTER 2

FUNDAMENTALS OF ACOUSTICS

Introduction. In both wire and radio telephony speech, music, and hearing are involved. The speech sounds containing the information to be transmitted often originate in enclosed spaces such as rooms, telephone booths, and radio studios. The acoustical characteristics of these enclosed spaces may influence greatly the speech sounds that operate a telephone transmitter or a radio microphone.

Transmitters and microphones must create electric signals that vary in accordance with the sounds to be transmitted. At the distant receiving station, the electric signals are reconverted to sound waves. Before transmitters, microphones, receivers, and loudspeakers can be considered, acoustics and the nature of the process of speaking and hearing must be understood.

The Nature of Sound.[1, 2] The science of sound is termed **acoustics.** *Objectively,* sound is a physical phenomenon consisting of a wave motion in air or other sound-transmitting media. *Subjectively,* sound is a sensation produced by an outside stimulation and occurring in the organs of hearing. For the purpose of this book, sounds will be grouped into speech sounds, musical sounds, and noises.

From the physical standpoint, sound consists of a series of condensations and rarefactions produced in the transmitting medium by a vibrating body. A vibrating body such as a tuning fork or a column of air is required to produce the waves, and some elastic medium such as air is necessary for the transmission of the sound waves. It is sometimes desirable to consider sound waves as traveling only through air. In the broadest sense, sound waves consist of a series of condensations and rarefactions existing in *any* medium.

Sound waves serve to transfer a portion of the energy of the vibrating body to the telephone transmitter, the radio microphone, or the ear. As the vibrating object moves, it displaces the air particles and causes condensations and rarefactions in the air. These particles transfer the condensations and rarefactions from one part of the air to another. The vibrations of the air particles are very small, and are quoted[1] by one author to be from 0.00000005 inch for barely audible sounds to 0.004 inch for loud sounds. The to-and-fro motion of the air particles from their position of rest occurs in a direction parallel to the progress

of the sound wave. Sound waves are therefore longitudinal as distinguished from transverse waves in which the displacement is at right angles to the direction of wave propagation. This motion is continued as long as the sound continues, that is, until the energy imparted by the vibrating body to the air particles is dissipated as heat or is transmitted away.

Simple sounds such as those produced by tuning forks are sinusoidal in nature. Steady-state complex (or non-sinusoidal) sounds can be analyzed into a fundamental and a series of harmonics. This is accomplished by having the sound wave actuate a high-quality microphone that in turn drives an oscillograph from which a photograph can be made and analyzed by the Fourier series methods. Or, the microphone can be connected to a wave analyzer, and the analysis made electrically. Panoramic viewing of a complex steady-state sound wave also is possible with special apparatus.

The Transmission of Sound. The sound energy that reaches and actuates the ear, a telephone transmitter, or a radio microphone is usually transferred from the vibrating source through the intervening air. There is, therefore, a certain (small) amount of energy in the air in the form of sound waves at any instant. If a sound source is in the open and continues to vibrate, energy is continually emitted by it. The sound waves travel outward until this energy is spread out and dissipated; then the sound becomes inaudible. Some of the energy is dissipated by heating the air.

If, however, the waves in traveling from the source strike an obstruction in their path such as an ordinary wall, most of the sound energy is **reflected.** When the sound waves generated in a room strike the rigid plastered walls, about 95 per cent of the sound energy will be reflected back into the room.

If the wall is massive and *not set vibrating as a whole,* part of the energy which is not reflected is **absorbed** by the surface layers of the wall. The portion not reflected or absorbed is **transmitted** as a very feeble wave motion in the material of the wall itself. These phenomena are shown in Fig. 1. If the source of sound is in air and not in rigid mechanical contact with the walls or floor of a room and if the walls are massive so that they will not be set in flexural vibration by the sound waves, then very little sound energy will be transmitted to other rooms in the building. If the vibrating body is in rigid contact with the walls or floor of the room, then vibrations will readily be transmitted to all parts of the building. These vibrations then will set up sound waves in the air in the adjoining rooms.

When sound waves originate in air, they are transmitted through an

obstructing medium such as a partition in three ways. *First,* sound may pass through the air spaces of a porous material. *Second,* the sound waves in the air may produce corresponding condensations and rarefactions in the material of the obstruction. This wave, which is very feeble, then proceeds through the material and produces sound waves in the air on the opposite side. *Third,* sound energy may be transmitted by setting the obstruction as a whole in motion. The vibrating obstruction then sets up sound waves on the opposite side.

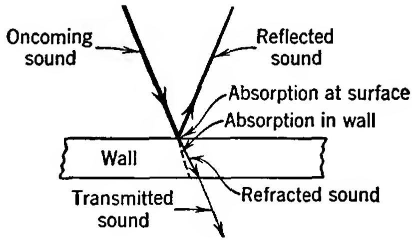


FIG. 1. Phenomena occurring when a sound wave strikes a rigid wall of some material having a velocity of transmission greater than air. The sound waves are shown as "rays" which represent the directions of travel of the wave fronts.

Sound absorption at the surface of a material is caused by the friction which occurs between vibrating air particles and the walls of the interstices in the surface of the absorbing material. The friction converts the wave energy into heat energy. If a wall or partition is caused to vibrate by sound waves, then internal friction between the fibers of the material will absorb sound energy. The transmission of sound through absorbing material varies according to the equation

$$i = i_0 \epsilon^{-ax}, \tag{1}$$

where i represents the sound intensity transmitted, i_o the sound intensity which is impressed, a is a constant depending on the type of material, and x is the thickness of the material. Measurements of the sound intensity sometimes are made with the Rayleigh disk resonator,[1, 3] but more often with a calibrated microphone (Chapter 4).

Sound Insulation. The important principles in sound insulation are, *first,* isolating the studio (or room) so that it is not in rigid mechanical contact with other portions of the building; *second,* constructing the walls, ceiling, and floor of tight, rigid material which will not permit sound "leakage" and which will not readily vibrate; and *third,* lining the walls with sound-absorbing material so that any sound which is transmitted into the room will be absorbed.

The isolation of machinery[4] and other noise-producing sources is important. Elaborate systems for isolating the walls and other boundaries of rooms have been designed.[5]

Echoes, Interference, and Beats. An **echo** is produced when a sound wave strikes an obstruction and is reflected. For example, a listener may hear a sound from a source first by direct transmission, and an instant later he may hear the sound by reflection from the walls. An echo becomes apparent to the listener only when the distance to the reflecting surface is such that the interval of time between the direct and the indirect receptions is about one-seventeenth of a second. Echoes may be very annoying in large auditoriums with curved reflecting surfaces which concentrate reflected sound. There are two methods of reducing echoes: *first,* by changing the form of the walls so that the reflected sound is scattered; and *second,* by covering the walls with sound-absorbing material so that the sound is absorbed and very little is reflected.

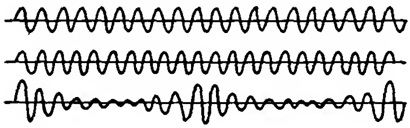


FIG. 2. Two pure tones of the same amplitude but of different frequencies combine as shown. The beat note which the ear "hears" is, however, created by the hearing mechanism itself.

Two or more sound waves from different sources tend either to increase or to decrease the sound intensity at different points, depending on whether the waves are related so as to add or to subtract at the point under consideration. Thus, if a steady tone of single frequency is sounded in a room, the waves directly from the source and those coming indirectly from reflecting surfaces such as walls will reinforce each other at certain points and will tend to neutralize each other at other points. At certain locations in the room, depending on the configuration of the room and the frequency of the sound, the sound will be intense, while at other points the sound will be weak.

When two sounds of different frequencies are simultaneously produced, the two waves combine instantaneously as in Fig. 2. It is often loosely stated that this phenomenon gives rise to a beat note. Such a beat note *is* created, but by the mechanism of hearing, and *not* in the air. The ear is a non-linear device, and such a device creates within itself new frequencies (in this instance the beat note) when two different frequencies are simultaneously impressed on it (page 414).

This can be proved in the following manner: Connect two good loud speakers to separate oscillators of good wave form. Set one oscillator on, for instance, 1500 cycles, and the other on 2000 cycles. The ear should then "hear" a 500-cycle tone. Next, pick up the resultant sound with a microphone and analyze the output with a

wave analyzer. This device will show that *only* a 1500-cycle tone and a 2000-cycle tone are present. If this is true, the 500-cycle "sound" must have been produced by the ear and brain, and *not* by the mere mixture or "beating together" of the two waves in the air.

These statements are not exactly correct for very intense sounds (reference 6, also, Fig. 23, page 44) but they do hold for the ordinarily observed phenomenon of beat notes.

Reverberation. If sound waves are produced in a room or other confined space, they remain audible after the source has ceased emitting until the energy in the waves is dissipated (largely in heat) and the level falls below that required for audibility. This prolonging of sound after the source has ceased emitting is termed **reverberation.** Both echoes and reverberation are caused by reflected sounds. They differ in that an echo is a "distinct" reflection often of an understandable sound, but reverberation is merely an indistinct "jumble" of reflected sounds. Excessive reverberation is the most common acoustical fault encountered in rooms, studios, and auditoriums. The practicable method of reducing reverberation is to add additional sound-absorbing material to the room.

In the past, many attempts have been made to correct the acoustic fault of excessive reverberation by the use of sounding boards and by stringing many feet of overhead wire supposedly to "break up" the sound waves and also to "carry" the sounds back to the rear of the auditorium. Although sounding boards can be used to advantage, there is no known scientific justification for the wires. Of course, what happens in many instances is that a large audience attends on the date of the "grand opening" to observe the effect of the wires, and their clothes absorb the sound, reducing the reverberation to a reasonable value.

W. C. Sabine placed architectural acoustics on a scientific basis.[7] He determined the sound-absorbing coefficients of a number of materials. Sound from a constant source builds up in a room until a state of equilibrium is reached when the sound is absorbed as rapidly as it is produced. The sound intensity will then become constant unless the energy from the source is changed. When the source is suddenly discontinued, the sound remains audible for a given period, depending on the absorbing power of the room.

Sabine, and later Jaeger, developed formulas[1, 7, 8] for the growth and decay of sound in a room. In 1929, Shuster and Waetzmann pointed out that these formulas did not hold for all rooms.[9] The equations now generally used are based on the work[9] of Eyring. These are, for

the growth of sound,

$$\rho = \rho_o\left(1 - \epsilon^{\frac{cS \log_\epsilon (1-\alpha)t}{4V}}\right), \tag{2}$$

and, for the decay of sound,

$$\rho = \rho_o\epsilon^{\frac{cS \log_\epsilon (1-\alpha)t}{4V}}. \tag{3}$$

In these expressions,

ρ is the sound energy per unit volume in the room at any time t.
ρ_o is the sound energy per unit volume when the sound has reached a steady state.
c is the velocity of sound in a room.
S is the total interior surface of the room.
α is the *average* value of the coefficient of absorption of the entire surface.
V is the volume of the room.

The sound energy in a room builds up and dies out according to the logarithmic curve of Fig. 3.

Sabine assumed that an open window was a perfect absorber of sound energy and that the absorption coefficient of an open window one foot square was unity, or, as now sometimes stated, one **sabine.** Other materials and objects used in constructing and furnishing rooms were rated with respect to an open window. In Table I a number of the more common materials are listed. For further information data published by manufacturers should be consulted; also, data will be found in handbooks.[5]

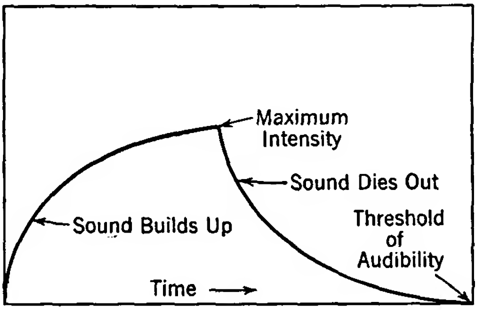


FIG. 3. Growth and decay of sound in a room.

It should be mentioned that the method of mounting affects the sound-absorbing properties of a material and that sound absorption is not entirely a surface phenomenon. If a material is mounted so that it is free to vibrate, under the influence of sound waves, internal friction between fibers of the material dissipates sound energy and hence affects the absorption coefficient. This fact is often overlooked.[10]

This diaphragm-absorbing action undoubtedly occurs in many mate-

TABLE I

Sound-absorption Coefficients of Typical Materials and Objects
(From *Electrical Engineers' Handbook*, Vol. 2, Electric Communication and Electronics, by H. Pender and K. McIlwain.)

Brand or Make	Thickness Inches	Sound-absorption Coefficients in Sabines at Various Frequencies in Cycles					
		128	256	512	1024	2048	4096
Sabanite (An acoustic plaster)	0.5	0.08	0.14	0.18	0.25	0.31	0.35
Absorbex, type A (An acoustic tile cemented to plaster, two coats of oil paint sprayed on)	1.0	0.07	0.22	0.51	0.91	0.77	
Acousti-Celotex, type B (A perforated acoustic tile, cemented to plaster board)	0.625	0.12	0.24	0.47	0.73	0.78	
Acousti-Celotex, type BBB (A perforated acoustic tile cemented to plaster board)	1.25	0.19	0.41	0.91	0.92	0.92	
Fir-Tex (An acoustic tile nailed to wood strips)	0.5	0.15	0.41	0.34	0.35	0.43	0.48
Glass		0.035		0.027		0.02	
Concrete, painted		0.009	0.011	0.014	0.016	0.017	0.018
Plaster	0.75	0.038	0.049	0.060	0.085	0.043	0.056
Wood Floor	0.75	0.09		0.08		0.10	
Carpets, lined		0.10		0.25		0.40	
Draperies, velour draped to $\frac{1}{2}$ area		0.14	0.35	0.55	0.72	0.70	0.65
Chair, plywood			0.19	0.24	0.39	0.38	
Chair, heavily upholstered theatre			3.4	3.0	3.3	3.6	
Person, adult fully dressed		1.8		4.2		5.5	

rials, even those designed to absorb energy by offering a rough, porous surface to the sound waves. If this were not true, painting unperforated acoustic material would reduce the absorbing characteristics to substantially those of painted wood. Tests and experience show that it definitely does not do so.

Measurements of Reverberation Time. As defined,[11] the **reverberation time** at a given frequency for an enclosure is the time required for the average energy density, initially in a steady state, to decrease along any simple or complicated decay curve to one-millionth

of its value when the source is cut off. The unit is the second. A decay of one-millionth of its initial value is 60 db (page 86).

Although the practice may not be satisfactory in all instances, it is common to regard reverberation time as the time it takes a sound which has reached maximum intensity in a room to decrease to inaudibility after the emitting source has been discontinued. This time may be approximately determined by loudly blowing some wind instrument such as a trombone or an organ pipe in a room and, after the sound has reached the maximum intensity, suddenly ceasing to blow and observing with a stopwatch the time it takes for the sound to become inaudible. Accurate instrumental methods of measuring reverberation time have been devised.*

Sabine early recognized that both the absorption coefficients of materials and the reverberation periods for rooms varied with the frequency used in performing the tests. The test frequency should accordingly be specified when such data are presented.

Reverberation Time in "Live" Rooms. Sabine developed the following approximate equation, which, although used for many years, was later shown to apply only to "live" rooms:

$$T = \frac{0.05V}{a}. \tag{4}$$

In this expression, T is the reverberation time in seconds, V is the volume of the room in cubic feet, and a is the total absorbing power of the surfaces exposed to the sound waves.

To illustrate the use of this equation, a typical calculation will be made. Room dimensions and other factors are assumed. The coefficients are from Table I at 512 cycles. A person seated, plus uncovered portion of the seat, is estimated at 5.0 units absorption.

Volume	200,000 cu. ft.		
Absorbing power			
Plaster	20,000 sq. ft.	at 0.06	= 1200 units
Wood	15,000 sq. ft.	at 0.08	= 1200 units
Vents	500 sq. ft.	at 0.75	= 375 units
Glass	1,000 sq. ft.	at 0.027	= 27 units
Upholstered seats	500	at 3.0	= 1500 units
Audience (seated)	500	at 5.0	= 2500 units
		Total	6802 units

$$\text{Reverberation } T = \frac{0.05 \times 200{,}000}{6802} = 1.47 \text{ seconds.}$$

* For information on this subject and on other aspects of reverberation, the *Journal of the Acoustical Society of America* should be consulted. References 12 and 13 also are recommended.

As will be shown later, 1.47 seconds is a satisfactory reverberation time for a room of this size *when occupied.*

Reverberation Time in "Dead" Rooms. With the advent of radio and sound-picture studios, it was recognized that Sabine's formula applied only to "live" as distinguished from "dead" rooms. A live room may be considered as one having a reverberation time of more than one second, and a dead room as one having a reverberation time of less than one second.

The reverberation time in *any* room is given[9] by the expression

$$T = \frac{kV}{-S \log_e (1 - \alpha_a)}, \tag{5}$$

in which T is the reverberation time in seconds, $k = 0.05$ (approximately, see reference 12) for a diffuse condition of the sound energy, V is the volume of the room in cubic feet, S is the total room surface in square feet, and α_a is the average absorbing coefficient. This average coefficient is found by the expression

$$\alpha_a = \frac{S_1\alpha_1 + S_2\alpha_2 + S_3\alpha_3 + \cdots}{S_1 + S_2 + S_3 + \cdots} = \frac{\Sigma S\alpha}{S}. \tag{6}$$

That is, α_a is found by adding the products of each different surface area and its corresponding absorption coefficient, and dividing this sum by the total room surface in square feet.

A simple correction makes equation 4 applicable to "dead" rooms. If $0.027V/S$ is subtracted from the reverberation time as given by equation 4, the corrected reverberation time will be substantially as found by equation 5. In this correction, V and S are as previously considered.[10]

Acoustics of Auditoriums. The reverberation time of an auditorium usually is the most important acoustical factor determining its suitability for both speech and music. The best reverberation time can be found from Fig. 4, which is based on curves given by several authorities.[10, 14, 15, 16] If the reverberation time is too long, then succeeding speech sounds will interfere and the words will tend to become indistinguishable. If the reverberation time is too short, then the speech sounds will be absorbed too quickly and the maximum intensity to which the sounds rise will be low. If an auditorium is to be used largely for music, a longer reverberation time is permissible than for speech, because music sounds best if the various tones are blended.

The curves of Fig. 4 are for a frequency of 512 cycles; but, as Table I shows, sound-absorbing properties vary with frequency. For

this reason, the reverberation time will be different at various frequencies. It is often satisfactory to use 512 cycles for auditorium design purposes. Curves are available[15, 16] for determining the best reverberation time at other frequencies.

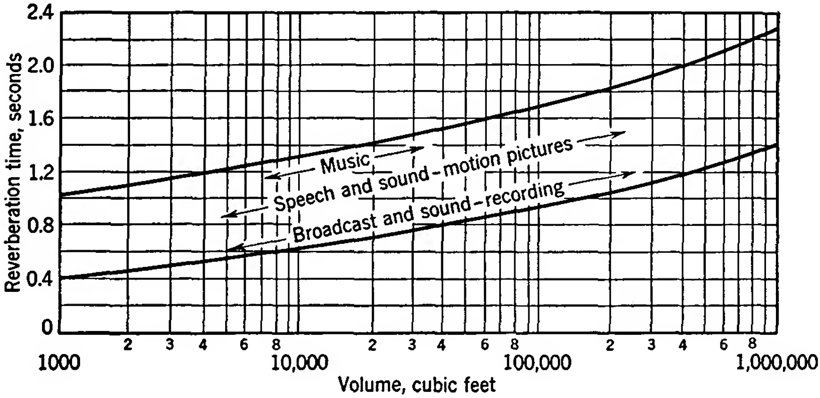


FIG. 4. Suitable reverberation times for rooms of various sizes and purposes. Frequency, 512 cycles. For instance, a small broadcast studio of about 1000 cubic feet should have a reverberation time of about 0.4 second. A large broadcast studio of 100,000 cubic feet should have a reverberation time of about 0.9 second. However, an auditorium of the same size used mainly for speech may have a reverberation time of 1.3 seconds; if used chiefly for music, it should have a reverberation time of about 1.6 seconds.

The practicable method of ensuring a correct reverberation time is to include sufficient sound-absorbing material. This should, of course, be installed when the auditorium is constructed, although it may be added later. For good results the shape of an auditorium should be more or less rectangular, without large curved surfaces.

The person speaking may be unable to supply sufficient sound energy to be heard satisfactorily in all parts of the auditorium. Accordingly, sound-amplifying equipment is installed in large auditoriums. Such equipment consists of a microphone, a vacuum-tube amplifier, and a system of loudspeakers. The *acoustic* output power requirements for such a system will depend on the size, reverberation time, extraneous noise level, and similar factors.[5] An appropriate idea of the acoustic power required in auditoriums of given size can be obtained from Fig. 5. If a sound-amplifying system is used it is permissible for an auditorium to have a lower reverberation time, and a lower reverberation time reduces acoustic feedback to the sound-amplifying system.

Acoustics of Studios. As indicated in Fig. 4, the reverberation time should be lower for studios than for auditoriums. It is advisable to design a studio so that the reverberation time will be correct over a frequency range from about 30 to 10,000 cycles. Such calculations necessitate a knowledge of the desirable reverberation time[15, 16, 17] and also a knowledge of the sound-absorbing qualities of the various acoustic and building materials over this frequency band.

Certain complex and intangible subjective effects such as tonal "brilliance," "brightness," and "liveness" must be considered in the design of broadcast and sound-recording studios.[15, 16, 17, 18] Studios

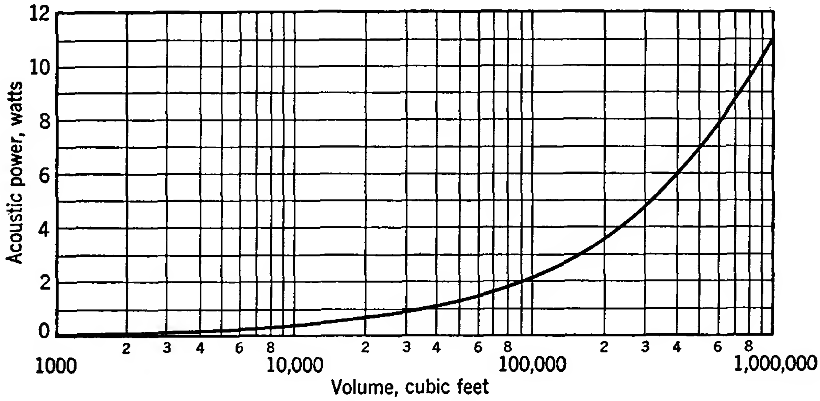


FIG. 5. Acoustic power required for full reproduction of speech and music in rooms of various sizes. This assumes optimum reverberation time and average noise conditions (Reference 5).

have been designed with "live" ends and "dead" ends. The live end contains little sound-absorbing material, and the musicians usually occupy this portion, so that the proper blending of the musical sounds is attained. The microphone is placed in the dead end, which is provided with much sound-absorbing material so that little reflection occurs. The modern tendency is not to concentrate the sound-absorbing material at a given area, but to spread it more or less throughout the room. Certain surfaces are often constructed of convex plywood "cylinders." These have absorbing properties at certain frequencies and also tend to vibrate and enhance certain frequencies.[17] Also, curved and non-parallel surfaces tend to prevent multiple reflection of certain frequencies from surface to surface that might result in "flutter."

Studios are usually approximately rectangular with the provision,

previously mentioned, that the surfaces are broken up with convex surfaces, or in other ways. To prevent the transmission of vibrations, and the entrance of sound from outside sources, studios are usually "floated" so that they are effectively isolated from the rest of the building.[5, 10] Air ducts for heating and ventilating must be carefully designed so that they do not convey sounds into the studio.

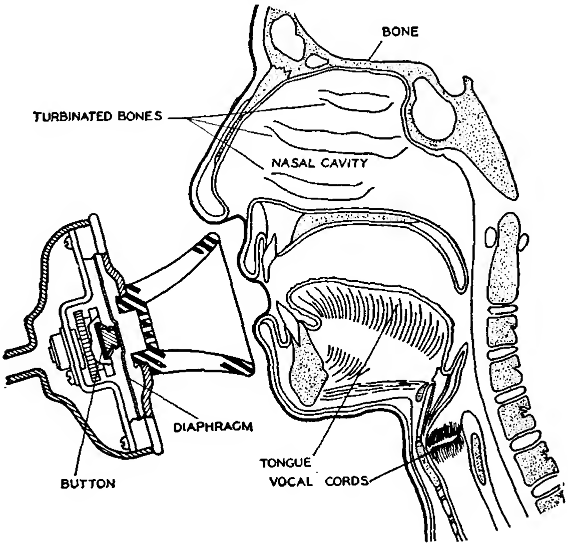


FIG. 6. Illustrating the resonant cavities provided by the throat, mouth, and nasal cavity. (Courtesy Bell Telephone Laboratories and D. Van Nostrand Company.)

The Production of Speech Sounds. All sounds of speech do not originate in the same manner. If the source of speech sounds is taken as a basic of classification, there are three different types: (1) those speech sounds produced by the vocal cords and the resonating air cavities of the head, (2) those produced by passing the air through small openings or over sharp edges in the mouth (for example, the teeth), and (3) the sounds produced by a combination of (1) and (2).

The human vocal organs consist of three principal parts. The first of these is the lungs and respiratory muscles, which provide a flow of air. The second is the larynx, containing the vocal cords as shown in Fig. 6, which produce the modulation of the air stream. The third part consists of the throat, mouth, and nasal cavities, which vary the harmonic content of the sound waves set up by the vocal cords, and which also originate certain sounds as mentioned in (2) of the preceding paragraph.

The vocal cords act as the generator for most speech sounds. They are capable of varying the shape and the width of the narrow slit which they form across the air passage.[19] In this way they control the pitch of the fundamental sound emitted. Under the influence of the stream of air passing between them from the lungs, these cords vibrate, producing sound waves. These waves are rich in harmonics, or overtones, and pass up into the resonant chambers provided by the throat, mouth, and nasal cavities. These are illustrated by Fig. 6. The size and shape of the mouth are varied by the speaker largely through movements of the lips and tongue, and certain resonant chambers are accordingly controlled. Through this action the waves from the vocal cords are given the characteristics of speech. The vocal cords, therefore, produce basic sounds which carry most of the energy of the voice waves but do not give speech its intelligibility. This is largely provided by the throat, mouth, and nasal cavities.

This explanation is substantiated by the artificial larynx. Through a surgical operation necessitated by disease, persons have had their larynges removed and have been left essentially speechless. The Bell Telephone Laboratories working in conjunction with surgeons devised an artificial larynx[20] which makes speech again possible for these unfortunate people.

For engineering purposes,[21] it is considered that there are certain fundamental speech sounds in spoken English. These are: (1) the **pure vowels,** (2) the **diphthongs,** (3) the **transitionals,** (4) the **semivowels,** (5) the **fricative consonants,** and (6) the **stop consonants.**

The Frequency Characteristics of Speech.[22] The fundamental pitch of the vowel sounds is subject to wide variations by changes in the vocal cords. The vowel sounds are characterized by resonant regions which are independent of the fundamental, and which determine the particular vowel sounded. Certain vowels, as shown in Fig. 7, are single-resonance vowels. Others have two regions of resonance, as Fig. 8 illustrates. The fundamental pitch in which the vowels are spoken has a decided influence on the spacing of these resonance points.[21] The characteristic frequencies of the semivowels and the nasal consonants are almost entirely below 3000 cycles. The stop and fricative consonants lie, for the most part, above 2000 cycles.

The complicated nature of speech sound waves and the electromagnetic waves used to transmit speech electrically can be seen from Fig. 9.

Although the pitch of the voice in speaking the vowels varies for different individuals, it is as low as about 90 cycles per second for a

deep-voiced man, and as high as about 300 cycles for a shrill-voiced woman. The harmonics, or overtones, go as high as about 6000 cycles.

In singing, the fundamental frequencies of the voice cover an approximate range from about 60 cycles for the lowest note of the bass

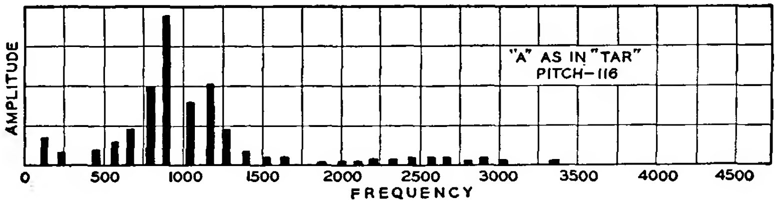


FIG. 7. A frequency analysis of a typical vowel having a single resonance region.

voice to 1300 cycles for the high notes of a soprano; the overtones go as high as about 10,000 cycles per second.

Extensive tests (page 33) have shown that it is not necessary to transmit all the frequencies present in speech or music sound waves.

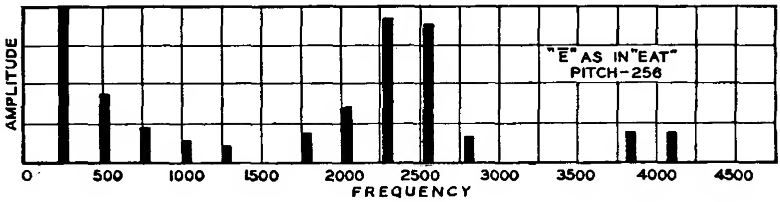


FIG. 8. A frequency analysis of a typical vowel having two resonance regions.

Speech Power. A number of speech power values have been defined.[21] Only instantaneous and average speech powers will be considered. **Instantaneous speech power** is the rate at which sound energy is being radiated at any instant, and **average speech power** is the total speech energy radiated over any period of time divided by the length of the period.

The amount of power in speech sounds is very small. To quote one authority,[21]

By taking the average speech power for a number of individuals talking in their usual conversational manner, it has been found that the average speech power for American speech is approximately 10 microwatts. If the silent intervals during conversation are excluded, this average is increased approximately 50 per cent. To carry this amount of power the air particles near the mouth vibrate through a distance of the order of 1/100 millimeter. When this amount

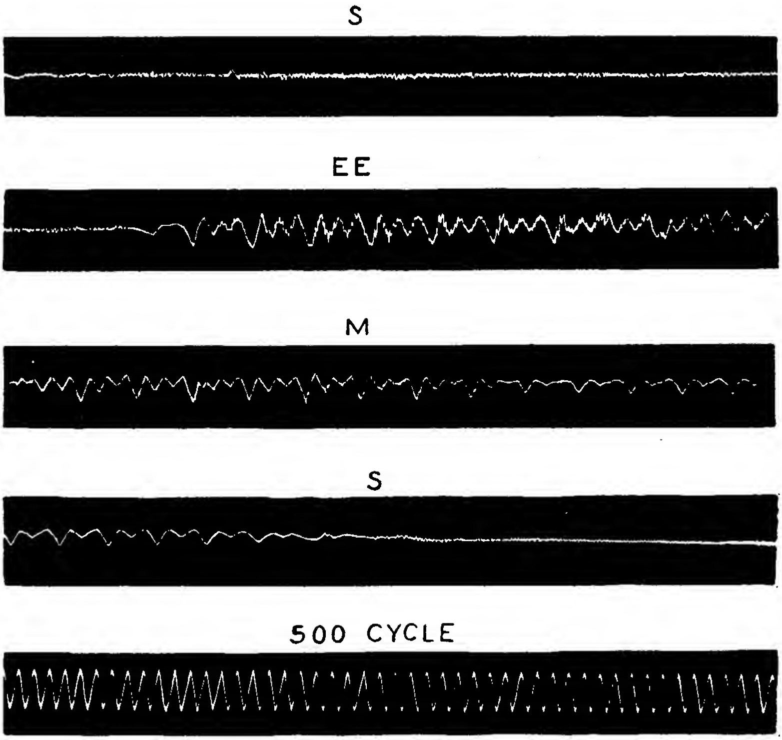


FIG. 9. Wave form of the word "seems."

of sound energy is received directly into the ear, it seems rather large due to the large excitation it produces on the auditory sense. However, it is really very small in comparison with the other powers ordinarily encountered. For example, it takes power equivalent to that produced by more than one million voices to light an ordinary incandescent lamp. It is therefore evident that the electrical currents used to transmit speech are of a different order of magnitude than those used to transmit power for lighting and heating purposes. It is only in some of the larger broadcasting stations that electrical speech currents are comparable in size with those used in power work.

When one talks about as loudly as possible the average speech power increases approximately to 1000 microwatts. When one talks in as weak a voice as possible without whispering it drops to 0.1 microwatt. A very soft whisper is about at 0.001 microwatt.*

The vowels are the most powerful speech sounds, having an average power value of about 100 microwatts, with peak values of about 2000

* Reprinted by permission of the publishers, from *Speech and Hearing*, by Harvey Fletcher, D. Van Nostrand Co.

microwatts. For any individual the ratio of the strongest to the weakest speech sounds is about 35 to 40 db (page 86). The energy distribution among the different frequencies[23] is shown in Fig. 10.

It is of interest to note that speech can be produced entirely by artificial means using electroacoustic equipment.[24] Such a device, called the *Voder*, was demonstrated at the 1939 World's Fairs in New York and San Francisco. Equipment for translating speech into visible patterns[25] has also been developed.

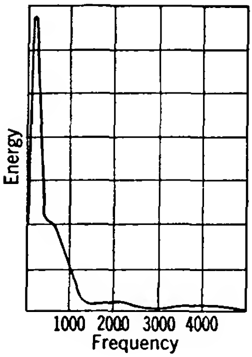


Fig. 10. Energy distribution; composite curve for male and female voices. (I. B. Crandall, *Bell System Tech. J.*, October, 1925.)

Music. The production of sounds by musical instruments depends on the action of the two principal parts of the instrument, the **generator** and the **resonator.** The function of the generator is to produce the tones. The resonator selects certain of these tones by resonant action and radiates them into the air. The vibrating string of a violin radiates very little sound directly. The amount radiated by the body of the instrument with which the string is in mechanical contact, however, is considerable.

Musical instruments may be classified into two major groups: first, those employing vibrating strings; and second, those having vibrating air columns. The piano and violin are examples of vibrating string instruments, and organ pipes and horns are typical of the second class. A third possible class of musical instrument is the percussion type of which the snare drum and cymbals are examples.

There are present, in any musical note, a large number of overtones, or harmonics. The frequency of the fundamental vibration determines the **pitch;** and the number, distribution, and relative intensity of the harmonics determine the **quality** of a note. It is this quality characteristic, depending as it does on the harmonics, which causes a note struck on a piano to be more pleasing and "richer" than a tone of the same fundamental frequency produced by a tuning fork or a vacuum-tube oscillator and loudspeaker. As an example, Fig. 11 shows both the wave form and the acoustic spectrum for a clarinet. The range of the fundamentals for musical instruments is as low as 16 cycles per second for an organ pipe, and as high as 4600 cycles for certain high-pitched instruments such as the piccolo.

The values just given are, as mentioned, for the fundamental frequencies only. The harmonics of musical instruments are of course

very much higher in frequency. Investigations[26] have shown that they go at least to 15,000 cycles (see Fig. 12).

In addition to a wide frequency range, music is characterized by wide changes in intensity. A 15-piece orchestra may have sound variations as great as 100,000 to 1 (50 db, page 86), during the rendi-

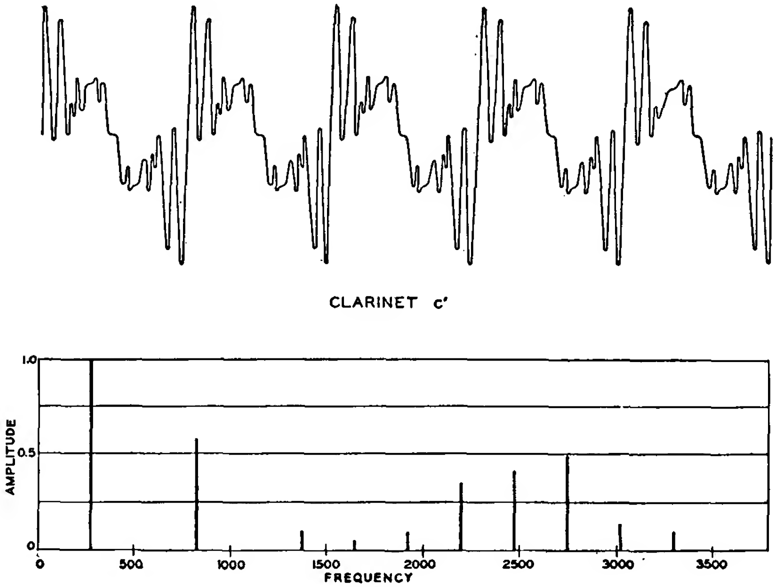


Fig. 11. Wave form and acoustic spectrum for c' (clarinet). (Courtesy Bell Telephone Laboratories and D. Van Nostrand Company.)

tion of a selection, and a very large orchestra will have even greater variations.

An extensive study[26, 27] was made to determine the frequency band and volume range that should be provided in the cable circuits for interconnecting radio broadcasting stations for musical programs. As a result of this investigation it was found that very little quality was lost if a band from 50 to 8000 cycles was transmitted, and if a volume range of 10,000 to 1 (40 db) was provided. In these tests the observers were trained musicians who noted when a *change* in the quality of the musical sounds was apparent.

A very interesting study was reported[28] in 1945. The observers were average radio listeners, and they recorded their *preference* when listening to narrow-, medium-, and wide-frequency-band transmission. The narrow-band system transmitted from about 120 to 4000 cycles,

the medium-band system from about 60 to 6000 cycles, and the wide-band system from about 30 to 10,000 cycles. As a result of this study the conclusions were drawn[28] that (1) the listeners preferred either a narrow or medium tonal range, and (2) most listeners still preferred a narrow to a wide tonal range, even when informed that one condition

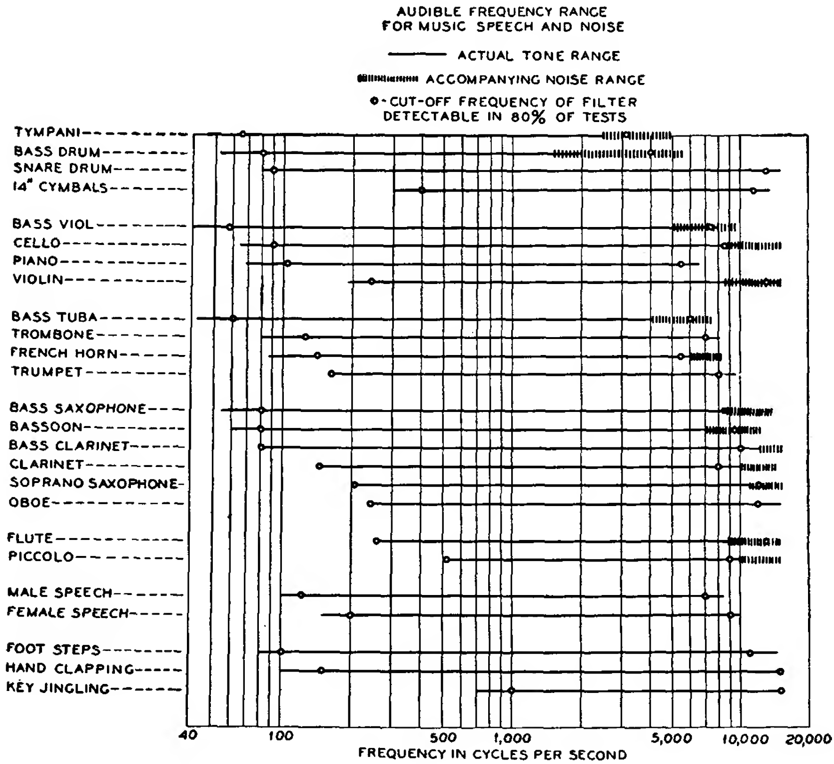


FIG. 12. (Reference 26.)

was low fidelity and that the other was high fidelity. This article has caused much discussion (*Proc. I.R.E.*, October, 1946, p. 757). A study with "live" talent instead of music from loudspeakers indicates that listeners prefer wide-band music. (See "Frequency Range Preference for Speech and Music" by H. F. Olson, *Electronics*, August, 1947.)

Noise. In communication, noises are objectionable because of their interfering effect. **Noise** is defined[11] as any unwanted sound. Thus, radio music may be undesired by a person using a telephone and is thus an interfering noise.

Noise, such as shown in Fig. 13, is perhaps the worst enemy of communication engineers. In telephone work, room noise affects the conversation in two ways. *First,* noises interfere with telephone conversation by the leakage of sound around the receiver cap. Noise has little interfering influence on the free ear of the telephone user. *Second,* room noise will actuate the telephone transmitter, and this in turn may operate the local receiver through sidetone (page 352) and also affect the distant receiver.

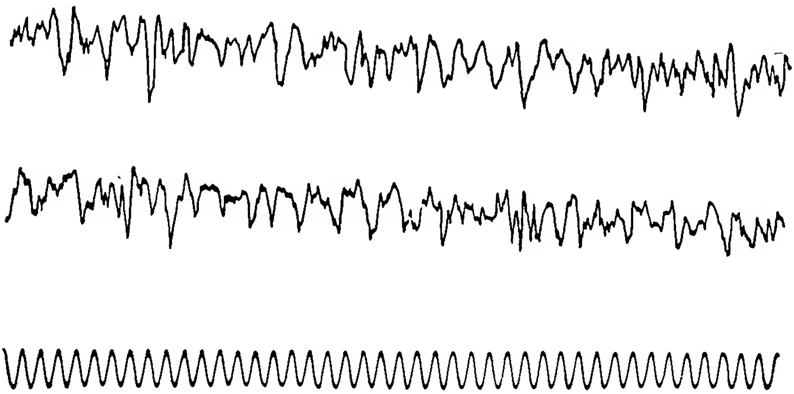


FIG. 13. Typical wave form of street noise and a pure tone of 500 cycles. (Courtesy Bell Telephone Laboratories and D. Van Nostrand Company.)

If the amount of noise cannot be reduced, an alternative is to raise the power level of the speech or music being transmitted. For example, troublesome static can often be made less noticeable if the power output of the broadcasting station is increased. The static as received is then relatively weak compared to the level of the program, and after amplification in the set is still relatively weak. Similarly, it is necessary to supply the average telephone user much more power than would be needed if the line and room noises were eliminated.

Hearing. The ear consists (Fig. 14) of the three following parts:

The Outer Ear. The external portion, for diverting sound energy into the ear, and the auditory canal, for conducting this energy into the eardrum, comprise this part of the ear.

The Middle Ear. The hammer is attached to the eardrum *D* and communicates to the anvil the motions imparted by the sound waves that strike the eardrum. These motions are then transferred by the anvil to the stirrup and through the foot plate to the oval window *O*, and thus to the inner ear.

The Inner Ear. There are three principal parts to the inner ear: namely, the semicircular canals, which serve only in maintaining equilibrium; the vestibule or space just behind the oval window; and the cochlea *S*, which serves as a terminating system for the mechanical sound vibrations, and in which these vibrations are converted into nerve impulses.

The cochlea is filled with fluid and is divided lengthwise into three parts by the basilar membrane and the membrane of Reissner. There are thus three parallel spiral canals, a cross section of which is shown

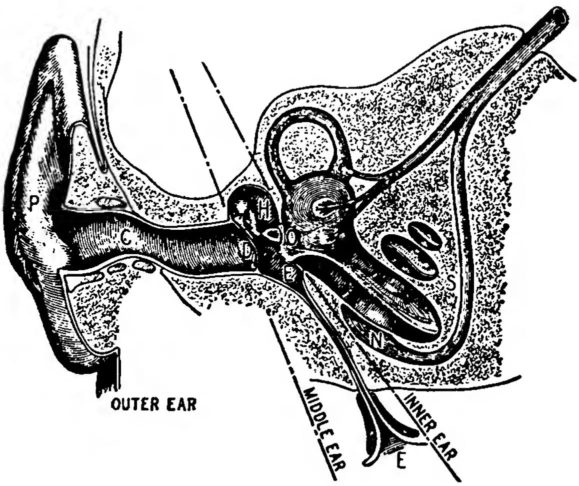


FIG. 14. Section of the human ear.

in Fig. 15. The scala vestibuli terminates at one end on the oval window *O*, and the scala tympani terminates on the round window *E*. These two canals are connected at the extreme end of the spiral.

The membrane of Reissner is very thin, and any impulses transmitted to the fluid in the scala vestibuli by the foot of the stirrup at the oval window are readily transmitted through this membrane to the fluid of the canal of cochlea. The flexible basilar membrane extends toward the upper end of the cochlea, dividing it as shown. An impulse transmitted to the fluid of the scala vestibuli will readily pass to the canal of cochlea and will set the basilar membrane in motion.

The organ of Corti containing the nerve terminals in the form of small hairs extending into the canal of cochlea is along one side of the basilar membrane. Lying over these small hairs is a soft, loose membrane called the tectorial membrane.

The process of hearing is as follows.[21] If a low-frequency note, below 20 cycles per second, impinges on the eardrum, this sound variation is transferred as a mechanical impulse to the fluid of the scala vestibuli. This fluid is in direct contact (at the far end) with that of the scala tympani, as previously mentioned. At this low frequency, the liquid offers little reactance, owing to its mass, and therefore the liquids in the two canals move bodily back and forth. The basilar membrane is not affected, and accordingly no sound sensation is produced.

If a 1000-cycle tone is impressed on the ear, however, the mass reactance of the fluid is great enough so that the liquid does not move back and forth, and the impulse is transmitted through the membrane

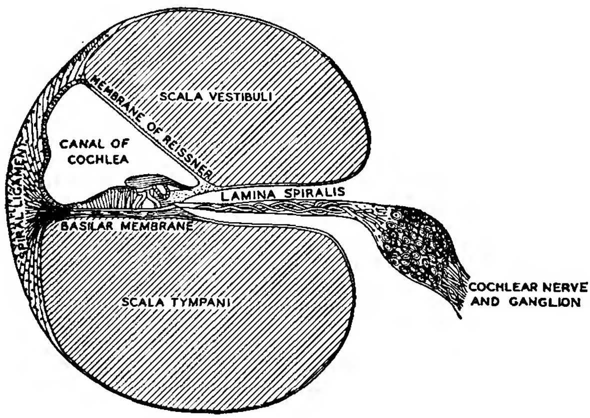


Fig. 15. (Courtesy Bell Telephone Laboratories and D. Van Nostrand Company.)

of Reissner to the canal of cochlea. The basilar membrane is caused to vibrate, and at some one spot the vibration will be greatest. The relative motion between the tectorial membrane and the basilar membrane then causes the small hairs to stimulate the nerve endings at their base, and this sends a sound sensation to the brain.

If a tone of higher frequency is used, a different part of the basilar membrane will vibrate with greatest amplitude, and thus different nerve endings will respond. When the frequency reaches and exceeds about 20,000 cycles per second, the hammer, anvil, stirrup, and associated parts absorb most of the sound energy, and little is transmitted to the inner ear; thus, the upper limit of audition results. It can be shown[21] that these three small bones together with the membranes on which they terminate act as a transformer to match the low impedance of the air to the high impedance of the fluid of the inner ear.

The Field of Audition. The field of audition is illustrated by Fig. 16. The data were obtained from tests of a large number of people with normal hearing. The lower curve indicates the intensity level necessary to produce a just audible tone at the different fre-

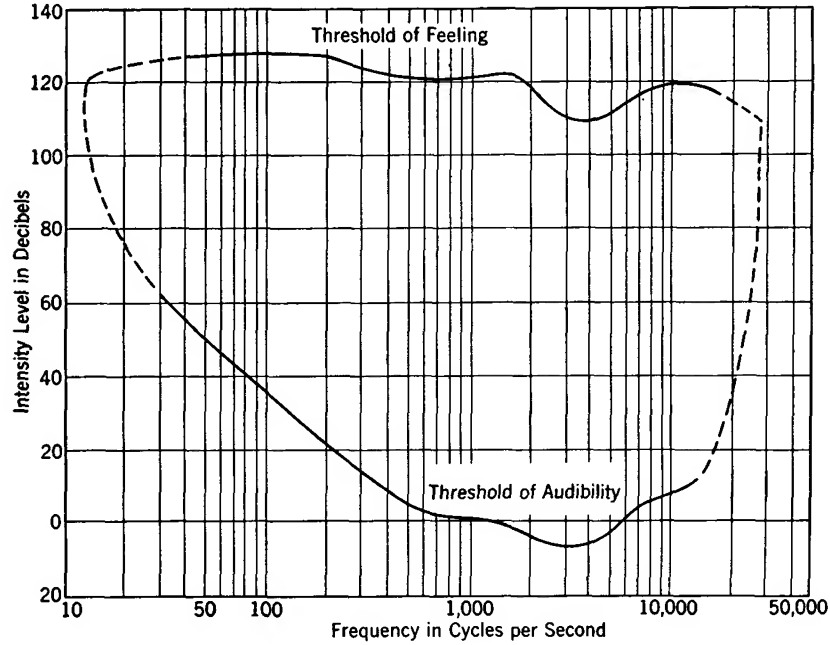


Fig. 16. Limits of audibility, or field of audition, of the human ear. Zero level represents a pure tone of 1000 cycles at an intensity of 10^{-16} watt per square centimeter. Actually, this threshold of audibility is for observers listening in a prescribed manner. Under these conditions, the hearing of the "reference observer" is slightly better than the average of a large group of individuals. See Reference 29. From a purely practical standpoint it is satisfactory to assume that the threshold of audibility for the average normal ear is 10^{-16} watt per square centimeter at 1000 cycles, although actually it is slightly higher.

quencies shown. The upper curve shows the intensity level for the different frequencies above which the impinging tone ceases to become a sound sensation and changes to a sensation of feeling.

A sound is audible if the frequency and intensity values lie within the area bounded by the curves. As is evident, the important hearing range for the average normal ear lies between about 20 and 20,000 cycles per second.

The Loudness of Sounds. The magnitude of the *sensation* produced in the brain is termed the **loudness** of a sound.[21] Although the

loudness of a sound is related to the intensity, the two are not the same. A sound that is loud for one person may not be loud for another. Also, two different sounds which produce equal intensities at the ear may not sound equally loud to the observer.

These phenomena are illustrated by the loudness-level contours of Fig. 17. Experiments have shown that loudness depends on both the *intensity* and the *frequency* of the sound; that is, the hearing mechanism is not equally sensitive to all intensities and to all frequencies. Loudness is the magnitude of the *sensation* produced in the brain. It

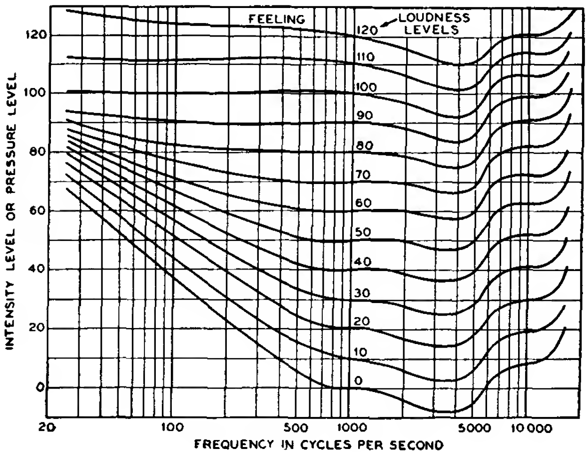


FIG. 17. Variations in the frequency response of the human ear with sound intensity. (Reference 29.)

is subjective, depending on the judgment of the individual. The *loudness level* of a sound, rather than the *loudness*, is accordingly measured. Before further study of Fig. 17 and the method of obtaining these curves, it is advisable to present several definitions largely summarized from reference 29.

Sound Intensity. The sound intensity of a sound field in a specified direction at a point is the sound energy transmitted per unit of time in the specified direction through a unit area normal to the direction at the point. The units for intensity commonly used are the erg per second per square centimeter or the watt per square centimeter.

Reference Intensity. The reference intensity for intensity-level comparisons shall be 10^{-16} watt per square centimeter.

Reference Tone. A plane or spherical sound wave having only a single frequency of 1000 cycles per second shall be used.

Intensity Level. The intensity level of a sound is the number of decibels above the reference level.

Loudness Level. The loudness level of any sound shall be the intensity level of the equally loud reference tone at the position where the listener's ear is to be placed.

Thus, in Fig. 17 the *loudness levels* of the various pure tones (rather than the loudness) are shown. For obtaining these data, observers, using special apparatus,[30] compared the sound to be measured with a 1000-cycle reference tone and adjusted the intensity level (in decibels above the chosen zero level of 10^{-16} watt per square centimeter) of the reference tone until the two sounds were heard as equally loud. Then, the intensity level of the reference tone above the reference intensity is a measure of the loudness level of the tone under test.

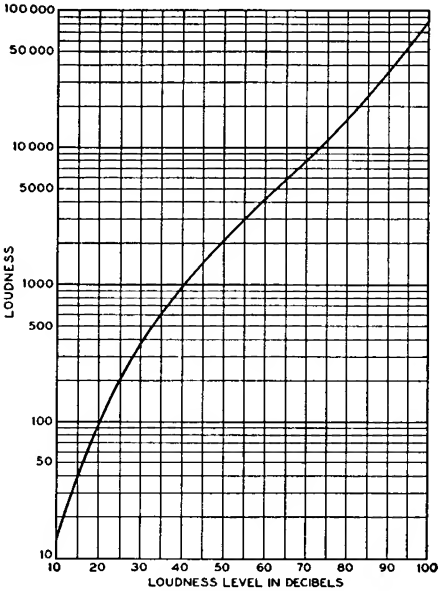


FIG. 18. The relation between loudness and loudness level. If the ear were exactly logarithmic the characteristic would be a straight line. (Reference 29.)

The curves of Fig. 17 are the characteristic curves of the human ear, showing its sensitiveness to sounds of different intensities and frequencies. Thus, at 100 cycles, a sound must be about 36 db above the reference intensity value at 1000 cycles to be just audible, but, at 10,000 cycles, it need be only about 8 db above the reference value to be audible.

Studies of the ear[21, 31] have shown that the increments of either energy or frequency that are necessary for the perception of differences in either loudness or pitch are logarithmic. The ear approximately follows the Weber-Fechner law of psychology, that sense perception varies as the logarithm of the stimulus. If the intensity level of the sound is changed 1 db (page 86), it will be found that this is about the smallest change that can be detected by the ear. The fact that equal perceptible changes in frequency are logarithmic is illus-

trated by the musical scale. The octaves are sensed as equal changes, although, from the frequency standpoint, the note of one octave is twice the frequency of the corresponding note of the next lower octave, and four times the frequency of the corresponding note two octaves lower.

The actual relation between loudness and loudness level is given by Fig. 18. This shows that the ear is not exactly logarithmic; if it were, the curve would be a straight line.[30, 32]

Masking. Telephones are often operated in noisy locations.[33] A study of the interfering effect of all the complex noises would be a large undertaking. The present discussion will be limited to the **masking effects** of pure tones.

In making these masking tests,[21] the *masking* tone was held at a constant intensity level, and the intensity level of the *masked* tone was gradually increased until it was just perceptible. This level in decibels to which the masked tone must be raised over that necessary without the presence of the masking tone is termed the **threshold shift.** Typical masking curves are shown in Fig. 19.

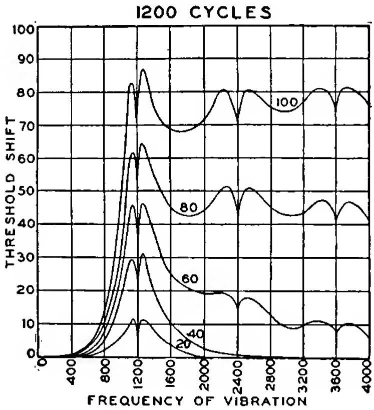


Fig. 19. Masking effect of pure tones. Masking tone 1200 cycles, at 20, 40, 60, 80, and 100 db above threshold value. (Courtesy Bell Telephone Laboratories and D. Van Nostrand Company.)

Noise Measurements. Noise causes a masking effect and a shift in the threshold of audibility. Thus, the level of a noise can be measured directly with the ear by determining the interfering effect. This is a *subjective* method. Or noise can be measured *objectively* by a suitably designed instrument.[34]

In the early noise surveys[35, 36, 37] both methods were employed. For the aural method, one arrangement of the apparatus is indicated in Fig. 20. The phonograph plays a special record producing a warbling or variable-frequency tone. The set is first calibrated in a sound-proof room for zero noise by adjusting the apparatus so that with no noise present the warble tone is just audible. Then, when the set is used to make noise measurements in a typical location, the attenuator is decreased until the warble tone is again just audible above the noise. The difference in the attenuator setting (measured in decibels)

is therefore equal to the deafening effect, or audibility threshold shift, of the noise.

Because of its practical advantages, the objective method of measuring noise has been generally adopted. Acceptable instruments for

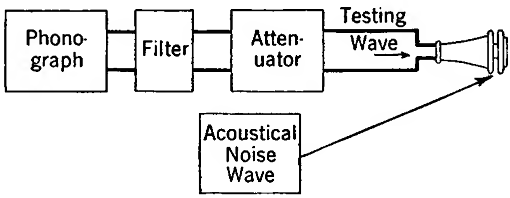


Fig. 20. Arrangement of apparatus for the aural method of measuring noise.

this purpose are designed following rigid specifications[38] based on the characteristics of the ear. A diagram of a typical sound- or noise-measuring instrument is shown in Fig. 21.

A high-quality microphone is used to intercept the noise to be measured, and its output is impressed on the preamplifier. A variable attenuator providing the various ranges is then included, after which

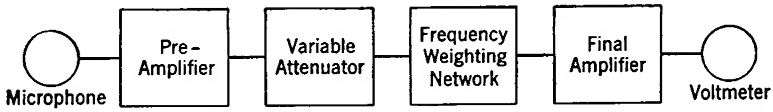


Fig. 21. Schematic diagram of a sound-level or noise meter.

the signal passes through a frequency-weighting network, a final amplifier, and to a suitable indicating instrument such as a vacuum-tube voltmeter.* Of course, provisions must be made for calibrating the circuit, either by using a standard tone, or by other means.[39]

The frequency-weighting network weights the various frequency components of a complex noise wave in accordance with the characteristics of the ear. The ear does not have the same characteristics at different loudness levels. Thus, for sounds of a level of 40 db, the

* With sound-level meters used in the United States the sound level in decibels above a zero level of 10^{-16} watt per square centimeter is obtained. By international agreement (see *Bell Laboratories Record,* February, 1938, page 214), the decibel is to be used for sound-intensity-level measurements, and the **phon** is to be used for loudness-level measurements. Thus, it is proposed that, if loudness-level measurements are being made as explained on page 40 and if the intensity level of the reference tone is n decibels, then the unknown sound has a loudness level of n phons.

ear has the characteristics given by the 40-db loudness-level contour of Fig. 17. For a 70-db sound, the characteristics are as given by curve 70, and for very intense sounds of about 100 db, the characteristics of the ear are essentially flat. Noise meters are designed with three different characteristics:[38] *first*, with a weighting network giving the correct characteristics for measuring weak noises (curve 40, Fig. 17); *second*, with a weighting network for noises of intermediate intensity (curve 70, Fig. 17); and *third*, a flat characteristic for intense noises and for general sound-measurement purposes.

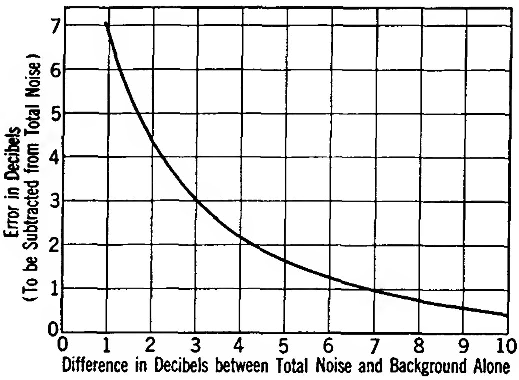


FIG. 22. Background noise correction for sound or noise-level measurements.

Background-Noise Correction. In industrial testing, and in other applications, the noise level of a machine or other device must be measured in the presence of background noises. Such noises would, of course, cause an error in the measurement, but these can be corrected very simply.[40]

The noise meter is placed in the test location, and the measurement is made of the background noise *without* the machine or device under test in operation. Then a measurement is made in the presence of the background noise *with* the machine under test operating. The correction to be subtracted from the total noise level is then determined from Fig. 22.

Noise Reduction. The sound-level, or noise, meter is useful for making quantitative studies of noise reduction. Noises produced within or transmitted into a room will remain audible until their sound energy is absorbed by the surfaces and objects within the room. In studying reverberation, it was shown that the intensity to which a sound (or noise) increases within a room is determined in part by the total amount of sound absorption present and is inversely

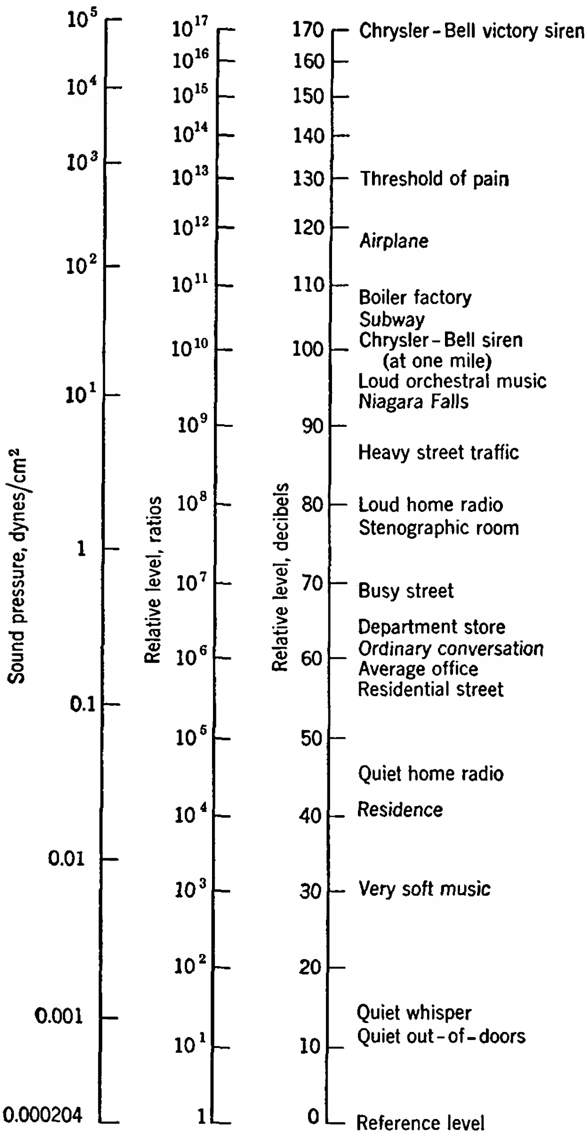


FIG. 23. Approximate pressures and levels of typical sounds. Zero reference level is 10^{-6} watt per square centimeter. Sound levels shown are close to source. The Victory Siren was audible at 75 miles under favorable conditions. The loudest possible *pure* tone is about 190 decibels. The loudest sound on record was produced by the explosion, in 1883, of the volcanic island Krakatoa in the East Indies. This sound was heard 3000 miles away. (Sirens and Bels, by J. O. Perrine, *Sci. Monthly*, February, 1943, Vol. LVI.)

proportional to it. If the sound output of the disturbing noise source remains constant, then doubling the absorption will halve the intensity, and so on. Thus, the noise level in a room can be reduced by the addition of sound-absorbing material.

The sound reduction in decibels is[10]

$$\text{Reduction in decibels} = 10 \log_{10} \frac{a_2}{a_1}, \tag{7}$$

where a_1 is the total units, or sabines, of sound absorption present, before treatment, and a_2 is the total amount after treatment.

From Fig. 18 it appears that a reduction in the loudness level of the sound of 65 to 60 db reduces the relative loudness from about 6000 to 4000, or to about two-thirds of its original amount. Thus, but a small reduction in the noise level in decibels causes a marked reduction in the interfering effect of the noise. Typical noise levels which have been found to exist are shown in Fig. 23. Useful information regarding noise and vibration studies is given in reference 41.

Articulation and Intelligibility Tests.[42] The problem of telephone communication is to convey intelligence in the form of the spoken word from one point to another. Tests of the ability of a telephone system to do this may be of two general types:[43] **articulation tests** and **intelligibility tests.** According to the reference just given, an articulation test measures the comparative reception of sounds *not conveying* ideas, and an intelligibility test measures the comparative perfection in the reception of sounds *conveying* ideas. These tests are based upon the general method[21, 44] of speaking into one end of the unit under test and having observers write out the sounds or sentences which they hear at the receiving end.

As a result of articulation tests valuable facts have been determined. One of these is the relation between the intensity of speech sounds and the articulation, the results of this study being shown in Fig. 24. Zero on this scale is taken[21] as the intensity level existing at the ear when a speaker talks in ordinary conversational tone with his lips ½ inch from the listener's ear.

Articulation is independent of the intensity level over a wide range of variations. As the curve shows, the speech intensity may be increased 100 times (or 20 db) or decreased to about one one-millionth (60 db) of the initial intensity without greatly affecting the articulation.

A study was made to determine the relation between articulation and frequency by inserting electrical filters (page 164) into the circuits being used. These filters were of two types—high-pass and low-pass

filters. The high-pass filters were designed to pass only those frequencies of the speech currents *above* a certain cutoff frequency; and the low-pass filters were designed to pass only those frequencies *below* a certain cutoff frequency. By inserting into the circuits different filters having different cutoff points and by making articulation tests the curves of Fig. 25 were obtained. "Articulation L" and "Energy L" were obtained with low-pass filters; the "H" curves were obtained with high-pass filters.

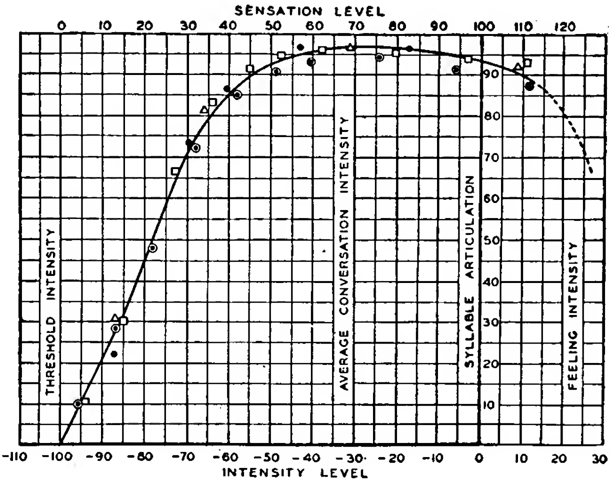


FIG. 24. Relation between the intensity of speech and articulation. (Courtesy Bell Telephone Laboratories and D. Van Nostrand Company.)

This figure shows that the lower frequencies carry most of the speech energy but that they contribute very little to the intelligibility. Also, these curves show that the high frequencies contain very little speech energy but contribute greatly to the articulation. If only those frequencies *above* 1000 cycles are allowed to pass through a circuit, the articulation is about 86 per cent perfect but the energy content is decreased to about 17 per cent of the total. If only those frequencies *below* 1000 cycles per second are transmitted, then the curves show that about 83 per cent of the total speech energy is present but that the articulation is reduced to about 42 per cent. This can be summed up in the statement that the high frequencies contribute most to the intelligibility, and the low frequencies contain most of the energy.

Commercial telephone service is provided if a band from about 250

to 2750 cycles is transmitted.[45] The frequency band has been widened to about 200 to 3500 cycles for many modern telephone circuits, giving better intelligibility and naturalness.

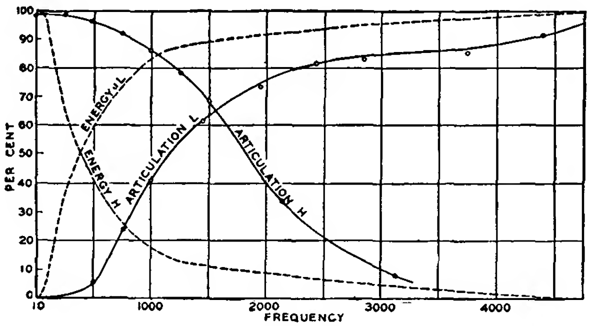


Fig. 25. The effects of eliminating certain frequency regions on the articulation and energy of speech.

The band-width requirements for broadcast program networks were discussed on page 33. Further information will be found in references 46 and 47. For frequency-modulation radio broadcast the band is arbitrarily set at 30 to 15,000 cycles.

REFERENCES

1. Watson, F. R. *Acoustics of Buildings.* John Wiley & Sons.
2. Sabine, P. E. *Architectural Acoustics.* McGraw-Hill Book Co.
3. Sivian, L. J. Rayleigh Disk Method for Measuring Sound-Intensities. *Phil. Mag.,* March, 1928.
4. Hansel, L. H. The Vibration Isolation of Machinery. *Elec. Eng.,* June, 1937, Vol. 56, No. 6.
5. Pender, H., and K. McIlwain. *Electrical Engineers' Handbook,* Vol. 2, Electric Communication and Electronics. John Wiley & Sons.
6. Thuras, A. L., R. T. Jenkins, and H. T. O'Neil. Extraneous Frequencies Generated in Air Carrying Intense Sound Waves. *J. Acoust. Soc. Am.,* January, 1935, Vol. VI.
7. Sabine, W. C. *Collected Papers on Acoustics.* Harvard Univ. Press.
8. Voorhees, L. E. Acoustics of Radio Broadcasting Studios. *J. A.I.E.E.,* March, 1930, Vol. 49, No. 3.
9. Eyring, C. F. Reverberation Time in "Dead" Rooms. *J. Acoust. Soc. Am.,* January, 1930, Vol. 1, No. 2, Part 1.
10. *Less Noise—Better Hearing.* The Celotex Corporation.
11. *Report of the Committee on Acoustical Standardization of the Acoustical Society of America.* Vol. 2, 1931.
12. Knudsen, V. O. *Architectural Acoustics.* John Wiley & Sons.
13. Olson, H. F., and F. Massa. *Applied Acoustics.* P. Blakiston's Son and Co.

14. MacNair, W. A. Optimum Reverberation Time for Auditoriums. *J. Acoust. Soc. Am.*, January, 1930, Vol. 1, No. 2, Part 1.
15. Content, E. J. Acoustical Design and Treatment for Speech Broadcast Studios. *Proc. I.R.E.*, February, 1944, Vol. 32, No. 2.
16. Content, E. J. The Acoustical Design of FM Studios. *Tele-Tech,* April, 1947, Vol. 6, No. 4.
17. Peterson, J. Acoustical Design of Broadcast Studios. *Tele-Tech,* March, 1947, Vol. 6, No. 3.
18. Jacobs, C. R. Acoustical Design of Studios for AM and FM. *Tele-Tech,* June, 1947, Vol. 6, No. 6.
19. Wegel, R. L. Theory of Vibration of the Larynx. *Bell System Tech. J.*, January, 1929, Vol. 9, No. 1.
20. Riesz, R. R. An Artificial Larynx. *J. Acoust. Soc. Am.*, January, 1930, Vol. 1, No. 2, Part 1.
21. Fletcher, Harvey. *Speech and Hearing.* D. Van Nostrand Co.
22. Hickman, C. N. An Acoustic Spectrometer. *J. Acoust. Soc. Am.*, October, 1934, Vol. 6.
23. Crandall, I. B., and D. Mackenzie. Analysis of Energy Distribution in Speech. *Phys. Rev.*, 19, 1922.
24. Dudley, H., R. R. Riesz, and S. S. A. Watkins. A Synthetic Speaker. *J. Franklin Inst.*, June, 1939, Vol. 227.
25. Potter, R. K. Visible Speech. *Bell Lab. Record,* January, 1946, Vol. 24, No. 1.
26. Snow, W. B. Audible Frequency Ranges of Music, Speech, and Noise. *J. Acoust. Soc. Am.*, July, 1931, Vol. 3.
27. Clark, A. B., and C. W. Green. Long Distance Cable Circuit for Program Transmission. *Bell System Tech. J.*, July, 1930, Vol. 9, No. 3. Also, *Trans. A.I.E.E.*, August, 1930, Vol. 49, No. 8.
28. Chinn, H. A., and P. Eisenberg. Tonal-Range and Sound-Intensity Preferences of Broadcast Listeners. *Proc. I.R.E.*, September, 1945, Vol. 33, No. 9.
29. *Standards for Noise Measurements.* American Standards Assn., publication Z24.2, February, 1936.
30. Fletcher, H., and W. A. Mundson. Loudness, Its Definition, Measurement, and Calculation. *J. Acoust. Soc. Am.*, October, 1933, Vol. 5.
31. Knudsen, V. O. The Sensibility of the Ear to Small Differences of Intensity and Frequency. *Phys. Rev.*, January, 1923, Vol. 21, No. 1.
32. Munson, W. A. Loudness Measurements. *Bell Lab. Record,* June, 1937, Vol. 15, No. 6.
33. Seacord, D. F. Room Noise in Telephone Locations. *Elec. Eng.*, June, 1939 and 1940, Vols. 58 and 59.
34. Churcher, B. G. Noise Measurements for Engineering Purposes. *Elec. Eng.*, January, 1935, Vol. 54.
35. Galt, R. H. Methods and Apparatus for Measuring the Noise Audiogram. *J. Acoust. Soc. Am.*, October, 1929, Vol. 1.
36. Galt, R. H. Results of Noise Surveys—Noise Out-of-Doors. *J. Acoust. Soc. Am.*, July, 1930, Vol. 2.
37. Williams, W. J., and R. G. McCurdy. A Survey of Room Noises in Telephone Locations. *Jour. A.I.E.E.*, September, 1930, Vol. 49.
38. *Standards for Sound Level Meters.* American Standards Assn., publication Z24.3, February, 1936.

39. Technique of Noise Measurement. *General Radio Co., Bulletin* 20.
40. *General Radio Experimenter,* December, 1937, Vol. 12.
41. JACK, W. A. Acoustics and the Quiet Train Ride. *Trans. A.I.E.E.*, 1942, Vol. 61.
42. CASTNER, T. G., and C. W. CARTER. Development in the Application of Articulation Testing. *Bell System Tech. J.*, July, 1933, Vol. 12.
43. COHEN, B. S. Apparatus Standards of Telephonic Transmission and the Technique of Testing Microphones and Receivers. *Trans. I.E.E. (London)*, Vol. 66, p. 165.
44. FLETCHER, H., and J. C. STEINBERG. Articulation Testing Methods. *Bell System Tech. J.*, October, 1929, Vol. 8, pp. 806–854.
45. MARTIN, W. H. Transmitted Frequency Range for Telephone Message Circuits. *Bell System Tech. J.*, July, 1930, Vol. 9, No. 3.
46. COWAN, F. A., R. G. MCCURDY, and I. E. LATTIMER. Engineering Requirements for Program Transmission Circuits. *Bell System Tech. J.*, April, 1941, Vol. 20, No. 2.
47. GANNETT, D. K., and I. KERNEY. The Discernibility of Changes in Program Band Width. *Bell System Tech. J.*, January, 1944, Vol. 23, No. 1.

REVIEW QUESTIONS

1. Name several important reasons why the fundamentals of acoustics should be understood by communication engineers.
2. Explain the differences between the objective and subjective interpretations of sound.
3. Explain how a continuous complex sound can be analyzed.
4. Discuss the phenomena that occur when a sound wave strikes a rigid wall. Repeat for a non-rigid wall that can vibrate.
5. If two sounds of slightly different frequencies and moderate intensities are simultaneously produced in air, is a beat frequency note created in the air? Explain the phenomenon that does occur.
6. What is meant by reverberation? What is the definition of reverberation time? How can a simple and approximate measurement of reverberation time be made?
7. What contributions were made by W. C. Sabine? Define the unit, sabine.
8. Why does the method of mounting a material affect its sound-absorbing properties? Does painting affect the sound-absorbing properties of a material?
9. What is the practicable method of reducing reverberation time? Why may the reverberation time be longer for music than for speech?
10. Enumerate the important considerations in the design of broadcast studios. Is a dead-air space between studio walls effective as sound insulation?
11. How are the sounds of speech produced?
12. Discuss and compare the frequency characteristics of speech and of music.
13. Discuss and compare the power in speech and musical sounds.
14. What studies have been made of the band-width requirements for radio programs, and what conclusions have been drawn?
15. Define noise. Why is it objectionable?
16. Describe the human ear and the process of hearing.
17. What is the field of audition, and what is its importance in communication?
18. Define loudness, and explain how it differs from sound intensity.
19. What is meant by a threshold shift?

20. Describe a sound-level meter.
21. In making noise measurements, how do you correct for background noise?
22. Discuss the theory of noise reduction.
23. Explain the difference between articulation and intelligibility tests.
24. What are the effects of removing the high-frequency components from speech, and from music?
25. What are the effects of removing the low-frequency components from speech, and from music?

PROBLEMS

1. Prepare a paper on the methods of measuring the sound-absorption coefficients of the various materials, furnishings, and objects used in radio studios.
2. Calculate the reverberation time for the auditorium considered on page 24 if plain plywood chairs are used, and with the auditorium empty, with 250 people, and with 500 people present. Would the auditorium be satisfactory for speech and music? If not, what would you recommend?
3. An auditorium is 125 feet long, 40 feet high, and 75 feet wide. The walls, ceiling, and floor are of glass, plaster, and painted wood. Assume that the average sound-absorption coefficient of these surfaces is 0.05 at 512 cycles. There are 1800 plain chairs in the room. Calculate the reverberation time when the room is empty, and with 500, 1000, 1500, and 2000 people present. Plot a curve showing these relations. How many square feet of sound-absorbing material having a coefficient of 0.47 should be used to reduce the reverberation time to an acceptable value when 900 people are present? Where should this material be placed?
4. One motor is mounted in each wing of an airplane. Analyze the means by which sound can be transmitted into the cabin, and explain the steps that should be taken to ensure that the cabin is reasonably quiet.
5. A room is 25 feet long, 15 feet wide, and 12 feet high. The walls and ceiling are plastered, and the floor is covered with linoleum. There are 10 desks, 10 chairs, and 10 persons in the room. Two of the persons are supervisors, and the remainder are typists and clerks. The noise level is 78 db. What should be done to reduce the level 5 db? To reduce it 10 db? What effect will this have on private conversations between the two supervisors, whose desks are adjacent?

CHAPTER 3

ELECTRICAL FUNDAMENTALS OF COMMUNICATION

Introduction. The currents, voltages, and powers used in communication range from microamperes, microvolts, and microwatts to many amperes, thousands of volts, and hundreds of kilowatts. The frequency range extends to billions of cycles. Communication systems consist of a large number of separate units, grouped together into complex electrical networks. The lines employed between distant cities are long, and phenomena that are not considered in other electrical work are of extreme importance in communication.

This chapter will be devoted to a brief consideration of those electrical phenomena that will later be of use in studying electrical communication. The material should be regarded as extremely essential, as many points of fundamental importance in communication will be presented.

Resistance. This is defined[1] as "the (scalar) property of an electric circuit or of any body that may be used as a part of an electric circuit which determines for a given current the rate at which electric energy is converted into heat or radiant energy and which has a value such that the product of the resistance and the square of the current gives the rate of conversion of energy."

Effective Resistance. This is measured[1] as "the quotient of the average rate of dissipation of electric energy during a cycle divided by the square of the effective current."

Resistance is given by the equation

$$R = \frac{P}{I^2}, \tag{1}$$

where R is the resistance in **ohms,** P is the power in watts, and I is the current in amperes.

When direct current is flowing in a circuit energy is dissipated in forcing the electrons through the wires. When alternating current is flowing through, there are *additional* losses. In communication cir-

cuits and apparatus, such as inductors, or coils, and capacitors, or condensers, the most important alternating-current and voltage losses are as follows:

Direct-Current Resistance Loss. One component of the alternating-current loss is the loss that would occur with direct current.

Skin-Effect Loss. When an alternating current flows through a wire, the rapidly changing magnetic field causes the current in the wire to crowd to the surface. This phenomenon is called **skin effect.** This reduces the cross-sectional area of the wire that is effective in carrying current, and the effective resistance is therefore greater than the direct-current resistance because of the **skin-effect loss.**

Magnetic Hysteresis Loss. Energy is required to reverse a magnetic field in a ferromagnetic material such as the core of a transformer. Thus, if an alternating current produces an alternating magnetic field in ferromagnetic material, a **magnetic hysteresis loss** occurs. The loss is directly proportional to the frequency.

Eddy-Current Loss. An alternating magnetic field induces voltages in objects in the vicinity, and, if the objects are conducting, **eddy currents** flow. These eddy currents will dissipate energy in the resistances of their paths and will cause an **eddy-current loss.** For eddy-current flow in transformer-core laminations the loss varies as the squares of the frequency and the thickness of the laminations.

Dielectric Hysteresis Loss. When an alternating voltage is impressed on a capacitor, an alternating electric field is established in the dielectric. Energy is dissipated in reversing this field. This is called a **dielectric hysteresis loss.** Such losses also occur in insulators, in the insulation of coils, and in all objects through which an alternating electric field passes. If the power loss per cycle is assumed to be constant, then the loss is directly proportional to the frequency.

Radiation Loss. Some energy is radiated by electric circuits, particularly at radio frequencies.

The important losses that occur in devices such as resistors, inductors, and capacitors have been enumerated. It is these additional losses that occur with alternating currents and voltages that cause the effective resistance to be greater than the direct-current resistance. It is common practice to use the term *resistance* when *effective resistance* is meant.

Self-Inductance. This is defined[1] as "the (scalar) property of an electric circuit which determines, for a given rate of change of current in the circuit, the electromotive force induced in the same circuit.

Thus

$$e_1 = -L\frac{di_1}{dt} \tag{2}$$

where e_1 and i_1 are in the same circuit, and L is the coefficient of self-inductance." In this equation e_1 is the instantaneous induced electromotive force in volts, L is the self-inductance in **henrys,** and di_1/dt is the instantaneous rate of change of current in amperes per second.

The *magnitude* of the instantaneous induced electromotive force is $e_1 = N\,d\phi_1/10^8dt$. If this is substituted in equation 2, the *magnitude* of the self-inductance becomes

$$L = \frac{N\,d\phi_1}{10^8\,di_1}. \tag{3}$$

Communication equipment such as coils and transformers often have closed cores of ferromagnetic material. Also, coils sometimes have direct-current and alternating-current components flowing simultaneously in the windings. The inductance of such units can be found by equation 3, assuming that the magnetic characteristics of the material are known. This is shown by Fig. 1 which illustrates the way in which the magnetic flux ϕ varies in a ferromagnetic core with the current i_1 that produces it. Suppose that the value of the direct-current components is i'_1 and that the alternating-current component is as shown by $\Delta i'_1$; then, the corresponding flux change will be $\Delta\phi'$. Hence, $d\phi_1/di_1$ of equation 3 will have a certain value, and the self-inductance L a certain magnitude. Now suppose that the value of the direct-current component i''_1 is different from i'_1, being larger as shown in Fig. 1. The same change in current $\Delta i''_1$ now produces but a very small change in magnetic flux $\Delta\phi''$, and hence $d\phi_1/di_1$ of equation 3 will

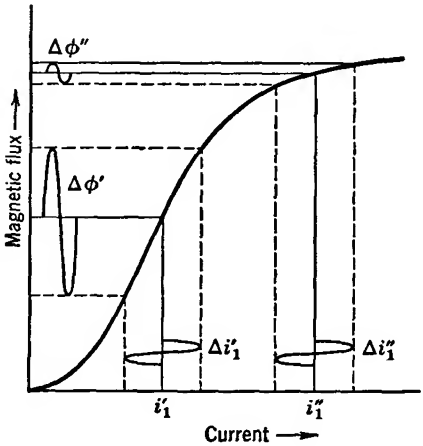


FIG. 1. Equal current changes $\Delta i'_1$ and $\Delta i''_1$ do not produce the same flux changes $\Delta\phi'$ and $\Delta\phi''$ if different values of direct current i'_1 and i''_1 flow through a coil with a closed core of iron or other ferromagnetic material. Thus, the incremental inductance depends on the value of the direct-current component.

be very small and the inductance will be much less than when the direct-current component is at i'_1. Because of these variations, the inductance of a coil on a ferromagnetic core is called the **incremental self-inductance,** or merely **incremental inductance.** This may be regarded as a **non-linear self-inductance.**

For coils and transformers with air cores, or ferromagnetic cores containing large air gaps, the inductance can be found by equation 3, or by the simplified form

$$L = \frac{N\,\phi_1}{10^8\,i_1}. \tag{4}$$

In such equipment the inductance is essentially independent of the magnitude of the current. The flux ϕ_1 produced by current i_1 is found by the usual methods. The self-inductance L will be in henrys when N is the number of turns, ϕ_1 is in lines, and i_1 is the current in amperes. This may be regarded as a **linear self-inductance.**

Sometimes inductance is measured with ordinary alternating-current measuring instruments that indicate effective values. For steady-state sinusoidal conditions and effective values, equation 3 becomes

$$L = \frac{E_L}{2\pi f I}, \quad \text{and} \quad E_L = 2\pi f L I, \quad \text{or} \quad E_L = \omega L I. \tag{5}$$

In these equations E_L is the magnitude of the effective value of the voltage drop in volts caused by the *inductive reactance* (excluding that caused by the effective resistance), I is the magnitude of the effective value of current in amperes, and f is the frequency in cycles per second. Such a determination would necessitate the use of a voltmeter, an ammeter, and a wattmeter to find the effective resistance or power-factor angle so that the reactive voltage drop E_L could be separated from the total voltage drop across the circuit or coil. The voltage E_L leads the current I by 90°. If the coil saturates and if the impressed voltage is sinusoidal, the current will be distorted and will contain harmonics (page 558). In this event, the ammeter will read the effective value of the current, and equation 5 will give **effective self-inductance.** A discussion of inductance when iron is present is given in reference 2.

Mutual Inductance. This is defined[1] as "the common property of two associated electric circuits which determines, for a given rate of change of current in one of the circuits, the electromotive force induced in the other. Thus,

$$e_1 = M\frac{di_2}{dt}, \quad \text{and} \quad e_2 = M\frac{di_1}{dt}, \tag{6}$$

where e_1 and i_1 are in circuit 1, e_2 and i_2 are in circuit 2, and M is the coefficient of mutual inductance." In these equations e is the instantaneous voltage in volts, M is the mutual inductance in **henrys,** and di/dt is the instantaneous rate of change of current in amperes per second. It is possible to write a form similar to that of equation 3.

$$M = \frac{N_2\, d\phi_1}{10^8\, di_1}, \tag{7}$$

For steady-state sinusoidal conditions and effective values, it can be shown that

$$M = \frac{E_s}{2\pi f I_p}, \quad \text{and} \quad E_s = 2\pi f M I_p, \quad \text{or} \quad E_s = \omega M I_p. \tag{8}$$

For convenience in applying to a transformer this equation has been written in terms of the effective value of the magnitude of the secondary voltage E_s induced by the effective value of the primary current I_p. The units are volts, amperes, and cycles per second. If phase relations are of importance a $-j$ should be placed in front of the final expressions of equation 8.

Capacitance. The definition of capacitance of most value in communication is[1] "the property of an electric system comprising insulated conductors and associated dielectrics which determines, for a given rate of change of potential differences between the conductors, the displacement currents in the system. Thus in a system of two conductors only

$$i = C\frac{de}{dt}, \tag{9}$$

where C is the capacitance, i the displacement current, and e the potential difference between the conductors." The usual units are amperes, farads, and volts per second.

Equation 9 is in terms of instantaneous values. For steady-state effective values this equation can be written

$$I = 2\pi f C E_C, \quad \text{and} \quad E_C = \frac{I}{2\pi f C}, \quad \text{or} \quad E_C = \frac{I}{\omega C}. \tag{10}$$

In these equations I is the magnitude of the effective value of the current in amperes, E_C is the magnitude of the effective value of voltage across the condenser in volts, and f is in cycles per second. The voltage E_C lags the current I by 90°. This equation assumes that capacitance only exists in the circuit.

Resistor. This is defined[1] as "a device, the primary purpose of which is to introduce resistance into an electric circuit." Unfortu-

nately, a rêsistor contains small amounts of "residual" inductance and capacitance. These unwanted properties become of extreme importance at radio frequencies. An equivalent circuit of a resistor is shown in Fig. 2.

Wire-Wound Resistors. Such resistors are used extensively in communication circuits, particularly where the frequency is low, where the current is relatively high, and where the resistors must be very stable and of the order of a few hundred thousand ohms and less.

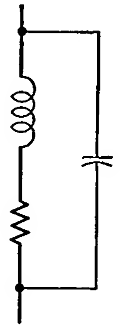


FIG. 2. An equivalent circuit of a resistor. The inductance and capacitance are undesired.

The methods of winding resistors to minimize inductance are such that the magnetic field produced is kept to the minimum by winding the resistance wire on a thin card, or by arranging the wires so that the flux-producing tendency of one wire is canceled by that of an adjacent wire.[3] Unwanted capacitances are minimized by reducing metallic areas and separating metallic portions between which a potential difference will exist.

Composition Resistors. This general heading includes several types of resistors. One is the so-called carbon resistor composed of a carbon composition formed into a short straight rod that is often encased in insulation. Another is the so-called metallized type consisting of a thin conducting film on an insulating rod, such as glass, and encased in insulation.

Composition resistors are particularly suited for radio because they are readily made with very high resistances, and because the residual capacitance and inductance are very low.

Inductor. This is defined[1] as "a device, the primary purpose of which is to introduce inductance into an electric circuit." There are **self inductors** and **mutual inductors,** but the dual wording is seldom used. An equivalent circuit of an inductor is shown in Fig. 3. The resistance shown is the effective resistance, and the capacitance is the **distributed capacitance** between turns. Inductors are often called inductance coils, choke coils, or merely coils and chokes.

Inductors used in communication are of two general types, those using air cores, and those using ferromagnetic cores, such as silicon steel laminations. For very low frequencies and where the core losses (eddy-current and magnetic hysteresis losses) are not of extreme importance, cores of silicon steel are used. When the losses must be kept to the minimum and high-quality performance is desired, cores of compressed powdered Permalloy (page 58) are employed. As

explained on page 52, the core losses will become very great at radio frequencies, and for this reason air cores are often used. There is a growing tendency, however, to use cores of compressed powdered ferromagnetic material such as iron at radio frequencies.

What is called the **energy storage factor Q** of an inductor is usually defined as the ratio of the energy stored to the energy dissipated per cycle. Numerically,

$$Q = \frac{X_L}{R} = \frac{\omega L}{R}. \tag{11}$$

This is the ratio of the inductive reactance of a coil to its resistance.

Capacitor. This is defined[1] as "a device, the primary purpose of which is to introduce capacitance into an electric circuit." The connecting leads and the metal foil and plates of a capacitor offer resistance and inductance to current flow; also, energy is dissipated in the dielectric. An equivalent circuit of a capacitor is as shown in Fig. 4.

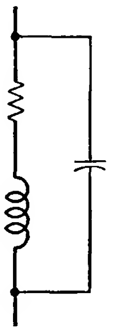


FIG. 3. An equivalent circuit of an inductor. The capacitance and resistance are undesired.

For direct-current, and at the lower communication frequencies and voltages, capacitors with metal-foil electrodes and wax or oil-impregnated paper dielectrics are used. Mica is used for the dielectric of capacitors where the losses must be kept low, or where high voltages are employed. Capacitors with ceramic dielectrics such as titanium dioxide and similar substances are used for small radio capacitors. Air is used as a dielectric for many radio capacitors, particularly if they are to be variable, must have very low loss, or are to be operated at high voltages. Electrolytic capacitors are used at low frequencies where a polarizing direct voltage is available, as in filters for power supplies.

FIG. 4. An equivalent series circuit of a capacitor. The resistance and inductance are undesired.

What is called the **energy dissipation factor D** of a capacitor is usually defined as the ratio of the energy dissipated to the energy stored per cycle. Numerically,

$$D = \frac{R}{X_C} = \omega RC. \tag{12}$$

This is the ratio of the equivalent series effective resistance to the capacitive reactance.

Magnetic Materials Used in Communication. The magnetic materials used in early telephone apparatus were chiefly magnetic iron and silicon steel. Bundles of small iron wire early were used for cores. About 1915 cores of compressed powdered electrolytic iron were introduced.

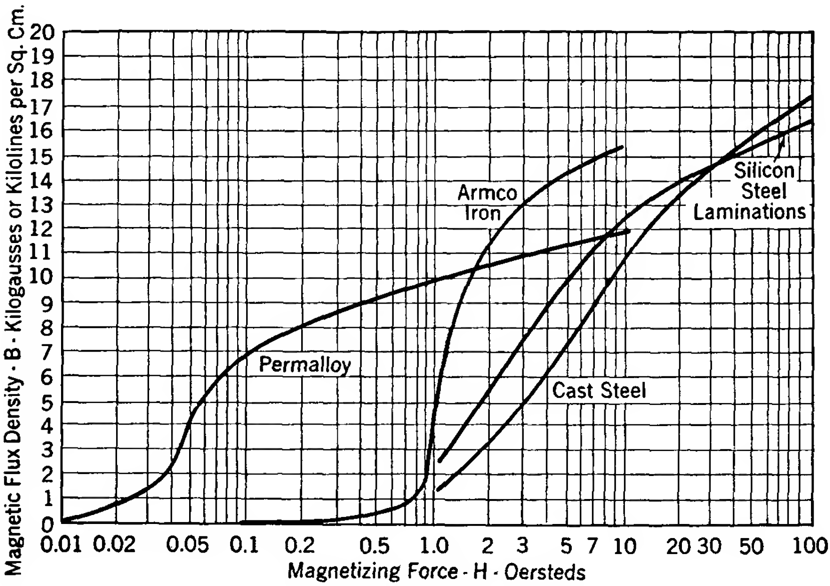


Fig. 5. Magnetization curves for certain magnetic materials.

Silicon-steel laminations are used in cores for power-supply transformers and for similar purposes. They are also used in some audio-frequency equipment. Special alloys have been developed for the specific requirements of telephone equipment, and certain of these have found use in radio.

Permalloys.[4, 5] This group of alloys contains varying amounts of nickel, iron, and chromium (Table I). The Permalloys have several characteristics of importance. *First,* they have high flux densities and permeabilities at the low currents encountered in communication apparatus (Figs. 5 and 6). *Second,* they have low hysteresis loss, as indicated by the small area of the hysteresis curve of Fig. 7. This keeps core losses low and also results in low non-linear distortion (pages 85 and 558). *Third,* the resistivity of the Permal-

loys is high (Table II), and hence the eddy-current losses are low. The eddy-current losses are minimized by making the cores of many devices of compressed powdered Permalloy. Each particle is effectively insulated from each of the other particles by a suitable insulating and binding material. Very thin Permalloy tape also is used in transformers. The Permalloys are used very extensively in the telephone industry. An improved Permalloy, called **Supermalloy** (su-perm′-alloy), has much higher initial and maximum permeabilities than the earlier Permalloys and has much less hysteresis loss.[6]

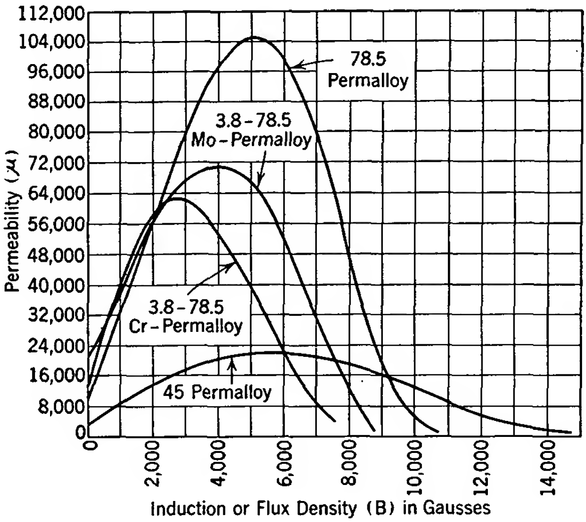


FIG. 6. Permeability curves for Permalloys. (Reference 4.)

Powdered Iron.[7, 8, 9] Special powdered iron cores have been used in radio-frequency coils and transformers since about 1935. Substances commonly used are[7] hydrogen-reduced iron, carbonyl iron, and magnetic iron oxide, commonly known as magnetite (Table III). These cores are usually molded "slugs" of finely divided iron, each particle of which is insulated and held by a suitable binder. In many radio coils and transformers the magnetic core is mounted on a screw so that the core may be moved inside the coil to vary the inductance. These coils are smaller, use less wire, have a higher Q, and are less expensive than comparable air-core coils.

Materials for Permanent Magnets.[10] Steels of high carbon content were used extensively for permanent magnets for many years. Later, cobalt steel, having better characteristics, was developed. More re-

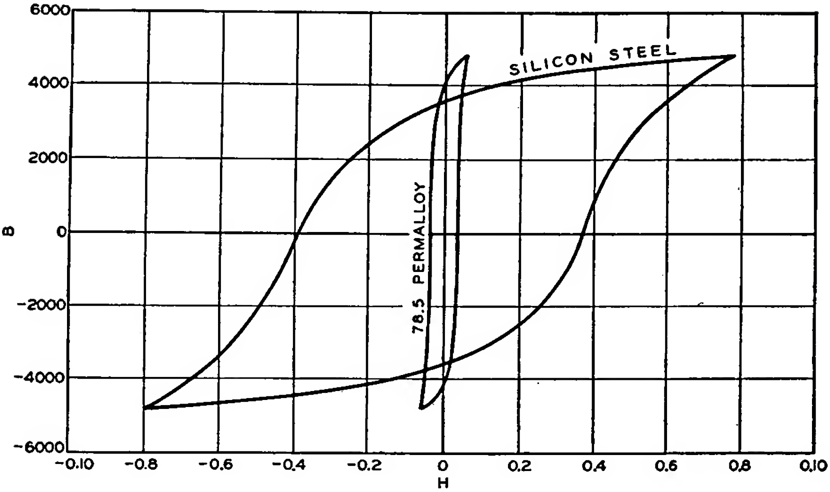


FIG. 7. Hysteresis loops of silicon steel and Permalloy. On the same scale, the loop for Perminvar would be a straight line.

cently, alloys (such as **Alnico**) of aluminum, nickel, iron, and other elements, such as cobalt, have been used for permanent magnets.

TABLE I

DESIGNATIONS AND COMPOSITIONS OF SOME COMMUNICATION MAGNETIC ALLOYS
(From reference 4.)

Designation	Composition, Per Cent					
	Ni	Fe	Co	Cr	Mo	V
78.5 Permalloy	78.5	21.5				
80 Permalloy	80	20				
45 Permalloy	45	55				
3.8–78.5 Cr-Permalloy	78.5	17.7		3.8		
3.8–78.5 Mo-Permalloy	78.5	17.7			3.8	
2–80 Mo-Permalloy	80	18			2	
45–25 Perminvar	45	30	25			
7–45–25 Mo-Perminvar	45	23	25		7	
Permendur		50	50			
1.7 V-Permendur		49.15	49.15			1.7

Ni = nickel; Fe = iron; Co = cobalt; Cr = chromium; Mo = molybdenum; V = vanadium.

A material called **Vicalloy** composed of vanadium, iron, and cobalt was announced[11] in 1940. The characteristics of certain of these materials are shown in Fig. 8.

TABLE II

Magnetic Constants for Communication Magnetic Alloys

(From reference 4.)

Material	μ_0	μ_m	$W_{H=\infty}$	B_r	H_c	$(B - H)_{H=\infty}$	ρ
"Armco" iron	250	7,000	5,000	13,000	1.0	22,000	11
4% silicon-steel	600	6,000	3,500	12,000	0.5	20,000	50
78.5 Permalloy, quenched	10,000	105,000	200	6,000	0.05	10,700	16
45 Permalloy	2,700	23,000	1,200	8,000	0.3	16,000	45
3.8–78.5 Cr-Permalloy	12,000	62,000	200	4,500	0.05	8,000	65
3.8–78.5 Mo-Permalloy	20,000	75,000	200	5,000	0.05	8,500	55
45–25 Perminvar, baked	400	2,000	2,500	3,000	1.2	15,500	19
7–45–25 Mo-Perminvar, baked	550	3,700	2,600	4,300	0.65	10,300	80
Permendur	700	7,900	6,000	14,000	1.0	24,000	6

Here μ_0 and μ_m are the initial and maximum permeabilities, respectively; $W_{H=\infty}$ is the hysteresis loss in ergs per cubic centimeter per cycle for saturation value of flux density; B_r is the residual induction in gausses; H_c is the coercive force in oersteds; $(B - H)_{H=\infty}$ is the saturation value of the intrinsic induction in gausses; ρ is the resistivity in microhms-centimeter.

TABLE III

Characteristics of Core Materials Used in Radio Coils

(From reference 9.)

Core	Hysteresis Loss Coefficient		Residual Loss Coefficient		Eddy-Current Loss Coefficient	
	μa	a	μc	c	μe	e
"70"	0.99	14.1	1.6	23	7.0	10.0
"55"	0.86	15.6	1.8	33	0.73	1.3
"40"	0.89	22.2	1.3	33	0.66	1.6
"40-L"*	1.45	37.2	2.7	69	1.00	2.5
"40-H"†	1.04	24.8	2.3	55	1.69	4.0
"16.5-E"‡	0.20	12	0.1	6	0.08	0.49

Quantities	Units
μa	10^{-3} ohms/henry, cps, gauss
a	10^{-6} ohms/henry, cps, gauss and per unit permeability
μc	10^{-2} ohms/henry, cps
c	10^{-5} ohms/henry, cps and per unit permeability
μe	10^{-7} ohms/henry, cps^2
e	10^{-9} ohms/henry, cps^2 and per unit permeability

* 40-L is a core made of Carbonyl iron, type L with a permeability of 39.

† 40-H is a core made of one of the best grades of commercial hydrogen-reduced iron, with a permeability of 42.

‡ Carbonyl iron type E with a permeability of 16.5.

Miscellaneous Magnetic Materials. Several other interesting magnetic materials have been developed for telephone purposes. Among these is **Perminvar**[12] listed in Table I. The permeability of this material is constant over a wide range of low flux densities, and the

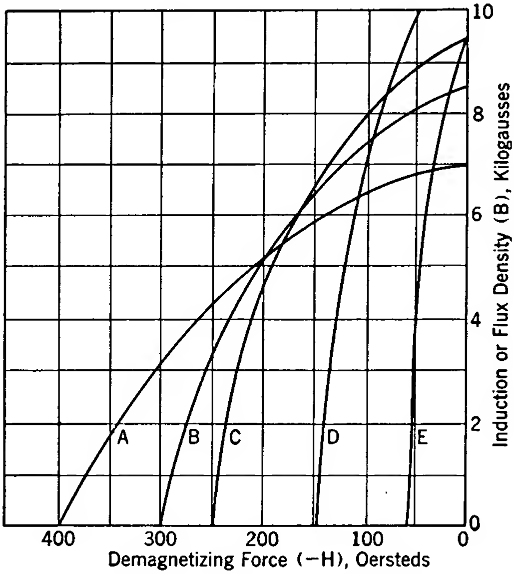


FIG. 8. Characteristics of certain permanent magnet materials. A large demagnetizing force required to reduce the flux to zero indicates a superior material. (Data from Reference 10.)

A. Nickel-aluminum-iron alloy. *C*. 35% Cobalt-steel alloy.
B. Cobalt-molybdenum-iron alloy. *D*. Cobalt-tungsten-iron alloy.
E. Chromium-steel alloy.

hysteresis loss is very low, even compared with Permalloy. Another group of alloys, known as **Permendurs,** have high permeability over a wide range of high flux densities.

Dielectric Materials Used in Communication. Materials used for insulators and dielectrics in communication apparatus are many and varied. A few of the common materials are listed in Table IV.

The power factors in Table IV indicate the dielectric losses in the materials. For paper dielectrics the dielectric constants range from about 3 to 5, and the power factors vary from about 0.1 to 0.5 per cent, depending on the nature of the paper-impregnating substance. Ceramic dielectrics such as titanium dioxide are used in radio circuits where constancy of capacitance with temperature change is desired and where the capacitors must be physically small.

TABLE IV

ELECTRICAL CHARACTERISTICS OF MATERIALS USED FOR INSULATORS AND DIELECTRICS

(From references 3 and 13.)

Material	Dielectric Constant	Power Factor — Per Cent 60 cycles	1000 cycles	10^6 cycles	Machine-ability
Cellulose acetate	6–8	7	. . .	3–6	Very good
Cellulose nitrate	4–7	5–9	5	5	Very good
Fiber	4–5	6–9	5	5	Very good
Glass, Pyrex	4.5	. . .	0.5	0.2	Very poor
Mica, clear India	7–7.3	0.03	0.02	0.02	. . .
Bakelite, "low loss"	5.3	2.5	1.4	0.7	Poor
Porcelain, "wet process"	6.2–7.5	2	1	0.7	Very poor
Rubber, hard	2–3	1	1	0.5–0.9	Fair
Steatite	6.1	1	0.4	0.3	Very poor
Styrene (polymerized)	2.4–2.9	0.02	0.02	0.03	Good
Titanium dioxide	90–170	. . .	0.1	0.06	. . .

Frequencies Used in Communication. An interesting characteristic of communication is the wide range of frequencies used. All designations have not been standardized.

Audio Frequencies. As explained on page 38, the average human ear functions on frequencies from about 20 to 20,000 cycles. The frequencies used in ordinary telegraphy, including teletypewriter operation, are below a few hundred cycles. The audio frequencies for *ordinary* commercial telephony extend from about 200 to 3000 cycles. For amplitude-modulation radio broadcasting the audio-frequency band usually extends from about 100 to 5000 cycles when network programs are used, and from perhaps 50 to 8000 cycles for local programs. For frequency-modulation broadcasting the audio band may extend from about 30 to 15,000 cycles.

Carrier Frequencies. In wire telephony and telegraphy the low-frequency signals are used to modulate a high-frequency carrier wave. This translates, or shifts, the information to be transmitted to the vicinity of the carrier frequency (page 412). Carrier frequencies range from about 4000 to 150,000 cycles, but in special systems frequencies as high as 500,000 cycles are used.

Radio Frequencies. Early radio systems operated at frequencies as low as about 10,000 cycles. The radio-frequency band extends from 10,000 cycles to many billions of cycles. The upper limit is being advanced as progress is made in radio. Radio frequencies are classified on page 442.

Series (Phase) Resonance. This is defined[1] as "the steady-state condition which exists in a circuit comprising inductance and ca-

pacitance connected in series, when the current in the circuit is in phase with the voltage across the circuit." From this definition the input impedance of the circuit at resonance is equivalent to pure resistance. This is often called **series resonance,** or merely **resonance.** The term **voltage resonance** is sometimes used, a term applicable only if the effective resistance is negligible.

Resonance occurs in a *series* circuit when the inductive reactance $X_L = 2\pi fL$ equals the capacitive reactance $X_C = 1/(2\pi fC)$. Thus,

$$2\pi fL = \frac{1}{2\pi fC}, \quad \text{and} \quad f = \frac{1}{2\pi\sqrt{LC}}, \tag{13}$$

where f is the (phase) resonant frequency in cycles per second when L is the inductance in henrys and C is the capacitance in farads. For the practical case a series circuit is composed of an inductor having effective resistance R and a capacitor having negligible loss. The resistance does *not* affect the frequency of resonance. The way in which the input impedance varies with frequency is shown in Fig. 9. Note that the input impedance is *low* at resonance.

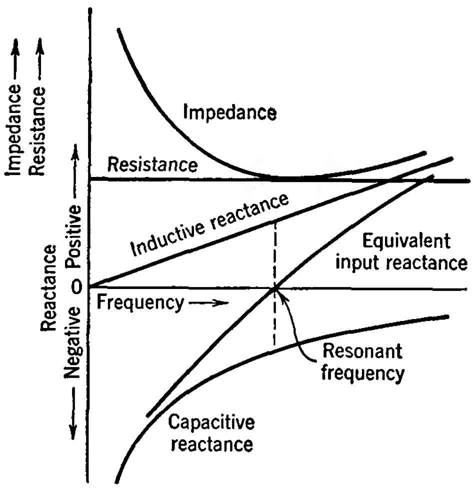


FIG. 9. Showing the variations in the resistance, reactances, and impedance of a series-resonant circuit. Note that at resonance the impedance equals the resistance. For a circuit of lower resistance (and higher Q) the impedance would fall to a lower value and the circuit would be more sharply tuned.

At resonance the current in a series-resonant circuit is limited only by the effective resistance of the coil. If the coil has a low effective resistance (a high Q), the current may rise to a high value, and large reactive voltages ($E_X = IX_C = IX_L$ approximately) will exist across the capacitor and inductor. This principle is used to increase signal voltages.

Parallel (Phase) Resonance. This is defined[1] as "the steady-state condition which exists in a circuit comprising inductance and capacitance connected in parallel, when the current entering the circuit from the supply line is in phase with the voltage across the circuit." From this definition the input impedance of the circuit at resonance is

equivalent to pure resistance. This is often called **parallel resonance.** The term **antiresonance** is sometimes used to designate parallel resonance. The term **current resonance** is also used, a term correctly applicable only if the effective resistance is negligible.

Resonance (phase) occurs in a *parallel* circuit composed of an inductor and capacitor when the inductive susceptance $B_L = X_L/Z_L^2$ equals the capacitive susceptance $B_C = X_C/Z_C^2$, and the reactive com-

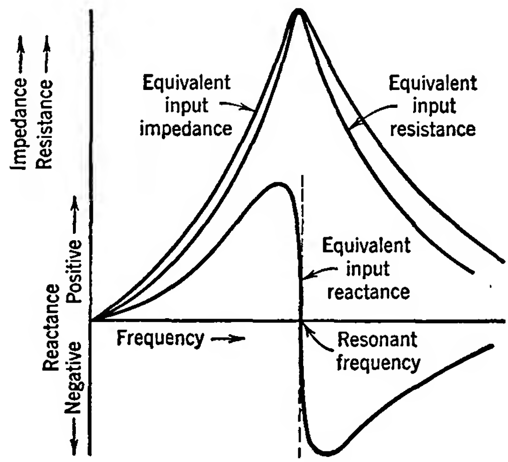


FIG. 10. Equivalent input impedance of a parallel-resonant circuit. Note that at resonance the impedance rises rapidly and becomes a large value of pure resistance. For a circuit with less loss (and higher Q) the curves rise to higher values and the circuit would be more sharply tuned.

ponent of the current through the inductor equals the reactive component of the current through the capacitor. For the practical case of a parallel circuit composed of an inductor having effective resistance R and a capacitor having negligible loss, the inductive susceptance is $B_L = X_L/(R_L^2 + X_L^2)$ and the capacitive susceptance is $B_C = 1/X_C$. The resonant frequency is

$$\frac{X_L}{R_L^2 + X_L^2} = \frac{1}{X_C}, \quad \text{and} \quad f = \frac{1}{2\pi}\sqrt{\frac{L - CR_L^2}{CL^2}}, \tag{14}$$

where f is the (phase) resonant frequency in cycles per second when L is the inductance in henrys, C is the capacitance in farads, and R_L is the effective resistance of the inductor in ohms. For many parallel circuits equation 13 is satisfactory. The way the input impedance of a parallel circuit varies with frequency is shown in Fig. 10. Note that the input impedance is *high* at resonance.

The voltage across a parallel circuit at resonance is $E = IR_e$, where

R_e is the equivalent parallel resistance. From the familiar equation for equivalent impedance of any parallel circuit,

$$Z_e = \frac{Z_1 Z_2}{Z_1 + Z_2}, \tag{15}$$

the equivalent parallel impedance of the inductor and capacitor previously considered is *approximately*

$$\begin{aligned} Z_e &= \frac{(R_L + jX_L)(-jX_C)}{(R_L + jX_L) + (-jX_C)} = \frac{(R_L + jX_L)(-jX_C)}{R_L} \\ &= \omega LQ = R_e. \end{aligned} \tag{16}$$

This last expression applies at radio frequencies where R_L is small compared with X_L, where R_L is neglected in the *numerator,* where $X_L = X_C$, and where $Q = \omega L/R_L$. The voltage at resonance across the parallel inductor and capacitor is $E = IZ_e = I\omega LQ$. Because the input impedance is greatest at resonance, the voltage across a tuned parallel circuit to which the current input I is constant also is greatest at resonance, and this property is made use of in practice to obtain a high signal voltage. At resonance the current through the inductor and capacitor are $I_L = E/Z_L$ and $I_C = E/X_C$, respectively, and may be quite large at resonance; this property also is utilized in practice.

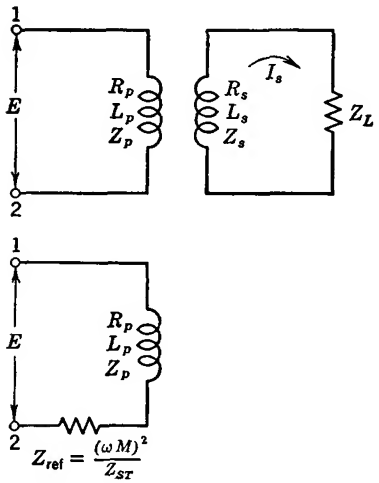


FIG. 11. The effect of a closed secondary circuit is to reflect an impedance Z_{ref} into the primary circuit.

Inductively Coupled Circuits. The current in the primary of a coil will induce a voltage in the secondary of the coil. If the secondary is open and no appreciable current flows through the secondary distributed capacitance, the effect of the secondary may be neglected. If the secondary is closed through some load impedance Z_L so that current does flow, then the effect of this current is felt in the primary, and the primary current is altered. Since the effect of a load connected to the secondary is to alter the current in the primary, the effect of the load is like connecting some impedance into the primary. This value of impedance is called a **reflected impedance** and is indicated in Fig. 11.

The magnitude and angle of the reflected impedance can be found as follows: Referring to equation 8, page 55, the voltage induced in the secondary will lag 90° behind the current in the primary, or $E_s = -j\omega MI_p$, and this must equal I_sZ_{st}, where Z_{st} is the total secondary impedance, including the impedance of the secondary of the coil and the load. Equating these two expressions, and solving for I_s gives

$$I_s = \frac{-j\omega MI_p}{Z_{st}}. \tag{17}$$

The voltage E must equal the sum of the voltage drop across the coil and voltage induced in the primary by the secondary, because this induced voltage must oppose flow of primary current. Thus,

$$\begin{aligned} E &= I_pZ_p + j\omega MI_s = I_pZ_p + j\omega M\frac{(-j\omega MI_p)}{Z_{st}} \\ &= I_pZ_p + \frac{(\omega M)^2I_p}{Z_{st}}. \end{aligned} \tag{18}$$

This equation can be written

$$E = I_p\left[Z_p + \frac{(\omega M)^2}{Z_{st}}\right] \tag{19}$$

which states that the effect of the closed secondary is to couple a load of impedance $(\omega M)^2/Z_{st}$ into the primary. All solutions must consider both the magnitudes and angles. To find the current flow in the secondary of Fig. 11 the steps are as follows:

1. Find the reflected impedance $(\omega M)^2/Z_{st}$, where $\omega = 2\pi f$, M is the mutual inductance in henrys, and Z_{st} is the *total* secondary impedance.
2. Find the total impedance between points 1–2, by adding the reflected impedance to the impedance of the primary of the coil.
3. Find the primary current by dividing the impressed voltage E (which may well be taken as the reference vector) by the total impedance between points 1–2.
4. Find the voltage induced in the secondary by the expression $E_s = -j\omega MI_p$.
5. Find the secondary current by dividing this induced voltage by the total secondary impedance.

Closely Coupled Circuits. When the primary and secondary of an inductively coupled circuit are on a closed ferromagnetic core a **closely coupled transformer,** or **transformer,** exists. In communication, transformers are often regarded as impedance changers, a viewpoint that will now be discussed.

Assume that a transformer, or **repeating coil,** as it is often called in telephony, has a closed core of Permalloy and that all the magnetic

flux ϕ produced by the primary winding N_p links with the secondary turns N_s; also, assume that the losses in the transformer are negligible. The magnitude of the back voltage induced in the primary is given by the fundamental expression $e_p = N_p\, d\phi/(10^8\, dt)$, and this will approximately equal the impressed voltage. The magnitude of the voltage induced in the secondary will be $e_s = N_s\, d\phi/(10^8\, dt)$. Dividing the second equation by the first, and in terms of effective instead of instantaneous values, gives

$$\frac{E_s}{E_p} = \frac{N_s}{N_p}, \quad \text{and hence} \quad \frac{I_s}{I_p} = \frac{N_p}{N_s}. \tag{20}$$

To determine the approximate impedance transforming equations, suppose that the transformer secondary is delivering power to a load. The magnitude of the load impedance will be $Z_L = E_s/I_s$, and the impedance measured across the primary terminals will be $Z_p = E_p/I_p$. From equation 20, $E_p = E_sN_p/N_s$, and $I_p = I_sN_s/N_p$, and, when these are substituted in the equation for the primary impedance,

$$Z_p = \frac{E_p}{I_p} = \frac{E_sN_p/N_s}{I_sN_s/N_p} = \frac{E_s}{I_s}\left(\frac{N_p}{N_s}\right)^2 = Z_s\left(\frac{N_p}{N_s}\right)^2. \tag{21}$$

This equation shows that the transformer acts like an impedance changer because the impedance Z_p measured at the primary will be $(N_p/N_s)^2$ times the impedance connected as a load to the secondary.

Maximum Power Transfer. A power system operates at constant voltage, but a communication circuit does not. The internal impedance of communication apparatus is often quite high, with the result that the output or terminal voltage varies greatly with the magnitude of the current taken by a connected load. The amount of electric signal power available from communication equipment (such as a telephone transmitter) is often very small, and maximum power transfer from one device or circuit to another, rather than efficiency of power transfer, is the criterion of good design.

The principle of maximum power transfer can be illustrated by a simple example. Suppose that a battery of constant open-circuit voltage E_{oc} and internal resistance R_i is connected directly to a load resistor R_L. It is desired to find the value of R_L such that the maximum power is transferred from the battery to the load.

The current flowing will be

$$I = \frac{E_{oc}}{R_i + R_L} \tag{22}$$

and the power P_L delivered to the load resistor R_L will be

$$P_L = I^2R_L = \frac{E_{oc}^2R_L}{(R_i + R_L)^2}, \tag{23}$$

where P_L will be in watts, when the other units are amperes, volts, and ohms.

If equations 22 and 23 are examined, the power P_L delivered to the load will be seen to approach zero if R_L approaches either zero or infinity in value. There is, accordingly, some intermediate point where the power transferred is maximum. This can be found by

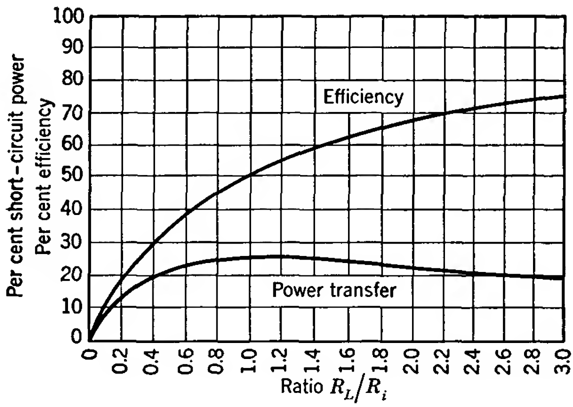


FIG. 12. Per cent of short-circuited power E_{oc}^2/R_i and efficiency for different values of R_L/R_i.

differentiating equation 23 and equating to zero, or by plotting a curve of equation 23. By either method it is found that *maximum power is transferred from the battery to the load when the load resistance equals the internal resistance of the battery.*

The relations just discussed are plotted in Fig. 12. The maximum power that can be developed by the battery is that produced when it is short circuited. The per cent of this delivered to the load is plotted against the ratio of the load resistance R_L to the battery internal resistance R_i, and maximum power transfer occurs when $R_L = R_i$. The efficiency of power transfer is

$$\text{Efficiency} = \frac{\text{power delivered}}{\text{power generated}} = \frac{I^2R_L}{I^2R_i + I^2R_L} = \frac{R_L}{R_i + R_L}, \tag{24}$$

and, when $R_L = R_i$, the efficiency is 50 per cent; that is, half of the power generated is lost in the battery, and half is delivered to the load.

In alternating-current circuits the power transferred from a generator of open-circuit voltage E_{oc} and internal impedance Z_g to a load of impedance Z_L will be $P = I^2R_L$, where

$$I = \frac{E_{oc}}{Z_g + Z_L} = \frac{E_{oc}}{(R_g + jX_g) + (R_L + jX_L)}, \quad \text{or}$$
$$\frac{E_{oc}}{(R_g + R_L) + j(X_g + X_L)}. \tag{25}$$

If the reactance X_L of the load is equal in magnitude and opposite in sign to the internal reactance X_g of the generator, then the reactances cancel, and the relations previously considered apply; that is, maximum power transfer occurs when $R_L = R_g$, and the efficiency is 50 per cent. Impedances, which have equal resistance components and reactance components equal in magnitude but opposite in sign, are defined[1] as **conjugate impedances.** For conjugate impedances the *maximum* power transferred becomes

$$P = I^2R_L = \frac{E_{oc}^2R_L}{(R_g + R_L)^2} = \frac{E_{oc}^2R_L}{4R_L^2} = \frac{E_{oc}^2}{4R_L}. \tag{26}$$

If the generator internal impedance and the load impedance are not conjugates, the power transferred is $P = I^2R_L$, where I is as given by equation 25. Thus, the power transferred is

$$P = \frac{E_{oc}^2R_L}{(R_g + R_L)^2 + (X_g + X_L)^2}. \tag{27}$$

When Z_g and Z_L are equal *both in magnitude and angle* (that is, $R_g = R_L$, and $X_g = X_L$), then equation 27 may be written

$$P = \frac{E_{oc}^2R_L}{4(R_L^2 + X_L^2)}, \quad \text{or} \quad P = \frac{E_{oc}^2R_g}{4(R_g^2 + X_g^2)}. \tag{28}$$

In communication circuits the following problem is often encountered: a load of impedance Z_L is to be connected to a generator of internal impedance Z_g. It is not possible to obtain a condition of conjugate impedances. Instead, the generator and load must be connected through a transformer that can be used to alter the *magnitude* only of the impedance. If the magnitude, but not the angle of an impedance can be altered, the relations for maximum power can be determined by rewriting equation 27 as follows

$$P = \frac{E_{oc}^2Z_L \cos \Theta_L}{(R_g + Z_L \cos \Theta_L)^2 + (X_g + Z_L \sin \Theta_L)^2}, \tag{29}$$

and by differentiating this expression and equating to zero.[14] The solution shows that *when the magnitude Z_L but not the angle Θ_L of a load impedance can be varied, maximum power will be transferred from a generator of internal impedance Z_g when the magnitude of Z_L equals the magnitude of Z_g.*

Impedance Transformations. The two preceding sections considered methods of changing the *magnitude* (but not the angle) of a load impedance. Transformers are often used for this purpose. Such methods are used in audio-frequency circuits where the band width is great compared to the magnitude of the center frequency. In contrast, the band width occupied by the radio channel is such a small per cent of the carrier frequency that in much radio design work it can be assumed that a single frequency is to be transmitted. Hence, it becomes possible to change angles as well as magnitudes in impedance matching. Such transformations will be treated in this section. The discussion applies to a single frequency, or to a relatively narrow band of frequencies.

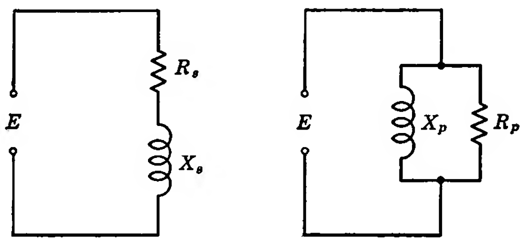


FIG. 13. Series and parallel circuits for studying impedance transformations.

Equivalent Series and Parallel Circuits. A method of making impedance transformations is based on equivalent circuits.[15] Two circuits are equivalent if they draw identical currents when they have the same voltage impressed on them. The current taken by the series circuit of Fig. 13 is $I_s = E/(R_s + jX_s)$, and by the parallel circuit $I_p = E/R_p + E/(jX_p)$. If the circuits are equivalent, these two expressions can be equated:

$$\frac{1}{R_s + jX_s} = \frac{1}{R_p} + \frac{1}{jX_p}. \tag{30}$$

For equivalence, the two in-phase components of current must be equal, and the two out-of-phase components must be equal. Thus in equation 30 the real terms must be equated separately, and the reactive

terms must be equated separately. When this is done,

$$R_p = R_s\left(1 + \frac{X_s^2}{R_s^2}\right) = R_s(1 + Q_s^2), \tag{31}$$

and

$$X_p = X_s\left(1 + \frac{R_s^2}{X_s^2}\right) = X_s\left(1 + \frac{1}{Q_s^2}\right). \tag{32}$$

These two equations are useful for finding the parallel circuit that is equivalent to a series circuit. In these equations $Q_s = X_s/R_s$ and is the Q of the *series* circuit that is to be converted into a parallel circuit.

Similarly, two equations can be written

$$R_s = \frac{R_p}{1 + Q_p^2}, \tag{33}$$

and

$$X_s = \frac{X_p}{1 + 1/Q_p^2}. \tag{34}$$

These two equations are useful for finding the series circuit that is equivalent to a parallel circuit. In these equations $Q_p = R_p/X_p$ and is the Q of the *parallel* circuit that is to be converted. Note that Q_p differs from Q_s.

Increasing an Impedance. Impedance transforming circuits are used extensively in radio. Usually the load to be matched contains both resistance and reactance. This reactance is often neutralized (and the phase angle of the load changed) by connecting an equal and opposite reactance in series. If this is done, then there remains only resistance to be transformed into a new value. A method of *increasing* a resistance will now be considered.

The circuit of Fig. 14(a) can be used to increase a resistance to a larger value. The resistance R may be a resistor or a circuit actually inserted in series with the inductor L, or it may be resistance reflected in series with the coil in accordance with coupled-circuit theory. Assume that R must appear as a larger value R' at the points indicated. The principle involved is to transform R to the desired value R' by connecting R in series with some inductor L. The equivalent inductance L' then is neutralized with capacitor C, leaving only the transformed value of R' effective between the terminals.

As an illustration, suppose that it is desired to transform 100 ohms resistance to 500 ohms resistance at 5.0 megacycles. Using equation 31

$$500 = 100\left[1 + \left(\frac{6.28 \times 5 \times 10^6 \times L}{100}\right)^2\right], \text{ and } L = 6.35 \times 10^{-6} \text{ henry.}$$

This is the value of inductance to insert *in series* with the resistance to be transformed. It is now necessary to find what the reactance of L' of Fig. 14(b) will appear to be. This can be found from equation 32, and is

$$2\pi f L' = X_p = 6.28 \times 5 \times 10^6 \times 6.35 \times 10^{-6}\left[1 + \left(\frac{100}{6.28 \times 5 \times 10^6 \times 6.35 \times 10^{-6}}\right)^2\right] = 250 \text{ ohms.}$$

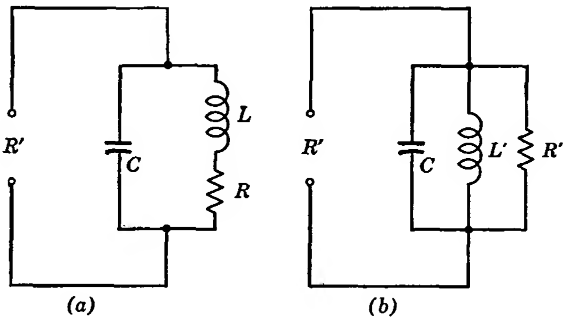


FIG. 14. Circuits for studying impedance transformations as used to increase a value of resistance R to a higher value R'.

The reactance capacitor C must have must also equal 250 ohms so that it will draw the correct leading current and neutralize the effect of the equivalent reactance of L'. Thus $1/(2\pi fC) = 250$, and $C = 127$ micromicrofarads. These calculations show that L of Fig. 14(a) should be 6.35×10^{-6} henry, and C of this figure should be 127 micromicrofarads to transform 100 ohms placed at R to 500 ohms as measured at R'.

Decreasing an Impedance. As an illustration, suppose that it is desired to transform a 500-ohm resistance to a 100-ohm resistance at a frequency of 10.0 megacycles. For this purpose, the circuit of Fig. 15(a) can be used. The method is to find the value of the inductor L that must be placed in parallel with the 500 ohms so that it will appear as 100 ohms at R'. Then it is necessary to find the equivalent series reactance of L and R so that this reactance can be neutralized with capacitor C, leaving only resistance R' between the input terminals. Using equation 33

$$100 = \frac{500}{1 + \left(\dfrac{500}{6.28 \times 10^7 \times L}\right)^2}, \quad \text{and} \quad L = 3.95 \times 10^{-6} \text{ henry.}$$

This is the value of the coil L to be connected *in parallel* with R to cause R' to equal 100 ohms. The equivalent series reactance of the L and R combination of Fig. 15(a) can be found by equation 34, and is

$$2\pi f L' = \frac{6.28 \times 10^7 \times 3.95 \times 10^{-6}}{1 + \dfrac{1}{\left(\dfrac{500}{6.28 \times 10^7 \times 3.95 \times 10^{-6}}\right)^2}} = 200 \text{ ohms.}$$

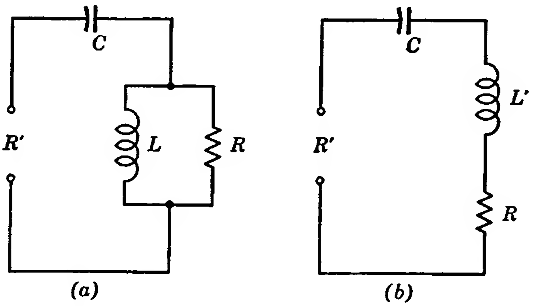


FIG. 15. Circuits for studying impedance transformations as used to decrease a value of resistance R to a lower value R'.

The capacitor C of Fig. 15 required to neutralize this inductive reactance is $200 = 1/(2\pi f C)$, and $C = 79.5$ micromicrofarads. Then, the resistance R' measured at the terminals of Fig. 15 will be 100 ohms.

Direct-Current Bridges. Measurements of direct-current resistance often are made with the Wheatstone bridge shown in Fig. 16. In operating, R_1 and R_2 are usually set on some convenient ratio, and resistor R_3 is then adjusted until the galvanometer does not deflect.

Since the galvanometer does not deflect, the points across which it is connected are at the same potential. Then,

$$I_1 R_1 = I_2 R_2 \quad \text{and} \quad I_1 R_3 = I_2 R_x.$$

Hence, $I_1/I_2 = R_2/R_1$, $I_1/I_2 = R_x/R_3$, and, when equated,

$$R_x = \frac{(R_2 R_3)}{R_1}. \tag{35}$$

This bridge is used to find faults such as short circuits, crosses, or grounds on communication circuits. One method is the **Murray loop** shown in Fig. 17. The lower wire on which the accidental ground exists is connected at the distant end to a clear wire. The bridge is then balanced.

Suppose that the two wires have the same resistance r per foot and that the total loop length L is known. Then, if A, B, Y, and X correspond to R_1, R_2, R_3, and R_x of Fig. 16 and equation 35, the *distance* X to the fault is

$$Xr = \frac{B(L - X)r}{A} \quad \text{and} \quad X = \frac{BL}{(A + B)}. \tag{36}$$

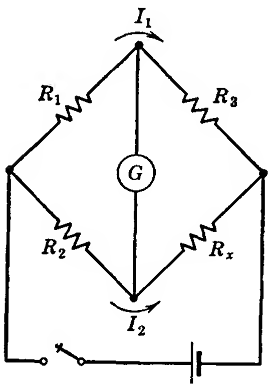


FIG. 16. The Wheatstone bridge.

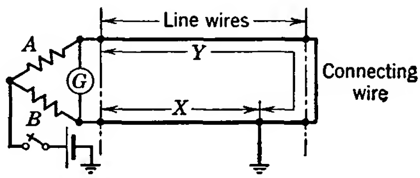


FIG. 17. The Murray loop.

If the two wires have different resistances r_1 and r_2 per foot because a wire of the same size was not available for connection with the faulty wire, $R_x = Xr_2$, and

$$R_3 = \frac{Lr_1}{2} + \left(\frac{L}{2} - X\right)r_2.$$

When these substitutions are made in equation 35, the distance is

$$X = \frac{BL(r_1 + r_2)}{2r_2(A + B)}. \tag{37}$$

Another method of finding faults is the **Varley loop** of Fig. 18. This arrangement offers the possibility of throwing the switch up and measuring the loop resistance. Suppose that in this figure the two wires have the same resistance r per foot and that the total loop resistance R_L has been measured. Then, when the switch is thrown down and the loop balanced, R_3 of equation 35 is $R_L - Xr$, and R_x is $R + Xr$. Hence, from equation 35,

$$R + Xr = \frac{B(R_L - Xr)}{A},$$

and the distance X to the fault is

$$X = \frac{BR_L - AR}{r(A + B)}. \tag{38}$$

In this equation, R_L is equal to the loop length L times the resistance per foot r. If the wires do not have the same resistance per foot, it is

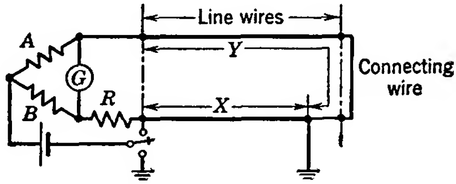


FIG. 18. The Varley loop.

easily shown that equation 38 can be used to find the distance to the fault if r is the resistance per foot of the grounded wire.

Alternating-Current Bridges.[16] These are used for measuring the effective resistance and inductance of inductors, the effective resistance and reactance of capacitors, and the input impedance of lines and networks. The condition of bridge balance usually is determined with a telephone receiver if the test frequency is audible. Above the audible-frequency range the null detector often is an amplifier and a rectifier. At higher frequencies, a radio-receiving set makes an excellent detector.[17, 18]

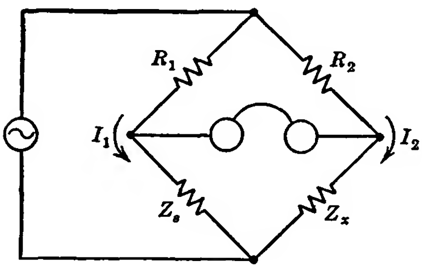


FIG. 19. An alternating-current bridge.

In the alternating-current bridge of Fig. 19 R_1 and R_2 are resistors with negligible reactance. Impedance Z_s may be a standard inductor or capacitor, and Z_x is the unknown inductor or capacitor to be measured.

At balance, negligible current flows through the receiver, minimum tone is heard in the receiver, and the points across which the receiver is connected are at the same potential. Thus $I_1R_1 = I_2R_2$ in *both* magnitude and phase, and $I_1Z_s = I_2Z_x$ in *both* magnitude and phase. Then,

$$\frac{I_1}{I_2} = \frac{R_2}{R_1}, \quad \frac{I_1}{I_2} = \frac{Z_x}{Z_s}, \quad \frac{R_2}{R_1} = \frac{Z_x}{Z_s}, \quad \text{and} \quad Z_x = \frac{R_2}{R_1} Z_s. \tag{39}$$

When this equation is generalized,

$$R_x + jX_x = \frac{R_2}{R_1}(R_s + jX_s) \cdot \tag{39a}$$

Because resistances and reactances produce different effects, in balancing an alternating-current bridge, first one quantity, and then the other, is varied, until minimum tone of the test frequency is heard. The in-phase and out-of-phase terms of equation 39*a* must be separated as follows: $R_xR_1 + jX_xR_1 = R_2R_s + jX_sR_2$, $R_xR_1 = R_2R_s$, and $X_xR_1 = X_sR_2$. Solving for the unknown terms gives

$$R_x = \frac{R_2}{R_1}R_s, \quad \text{and} \quad X_x = \frac{R_2}{R_1}X_s, \tag{40}$$

where R_x is the effective resistance and X_x is the reactance of the unknown. When a standard inductor is used to measure an unknown inductor, equation 40 becomes

$$2\pi f L_x = \frac{R_2}{R_1}(2\pi f L_s), \quad \text{and} \quad L_x = \frac{R_2}{R_1}L_s \cdot \tag{41}$$

When a standard capacitor is used to measure an unknown capacitor, equation 40 becomes

$$\frac{1}{2\pi f C_x} = \frac{R_2}{R_1}\left(\frac{1}{2\pi f C_s}\right), \quad \text{and} \quad C_x = \frac{R_1}{R_2}C_s, \tag{42}$$

in which the ratio R_1/R_2 is opposite from equation 41. In the preceding equations the units are ohms, henrys, farads, and cycles per second.

Bridge with Standard Inductor. The unknown inductor or capacitor of Fig. 20 is measured in terms of the standard inductor L_s and standard resistor R_s. The effective resistance of the standard inductor necessitates a correction. If Fig. 20(*a*) is used, then R_s of equation 40 equals R_s of Fig. 20(*a*) plus the effective resistance of the standard inductor. If Fig. 20(*b*) is used, then the effective resistance of the unknown capacitor is R_x of equation 40 minus the effective resistance of the standard inductor.

Often, in bridges, such as Figs. 20(*a*) and (*b*), R_1 and R_2 each equal some value, such as 1000 ohms, and are fixed. Also, a resistor, equal to the effective resistance of the standard inductor, is connected in the arm opposite the standard variable inductor. The bridge of Fig. 20(*a*) then becomes direct reading; at balance the setting of R_s gives the effective resistance of the unknown inductor, and the setting of L_s gives the inductance of the unknown.

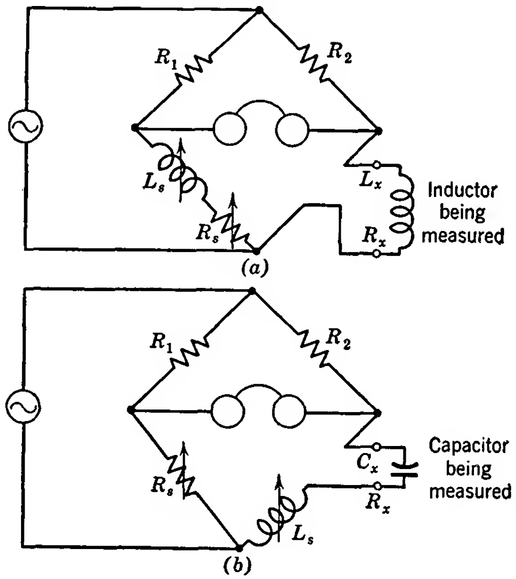


FIG. 20. An alternating-current bridge using standard inductor L_s.

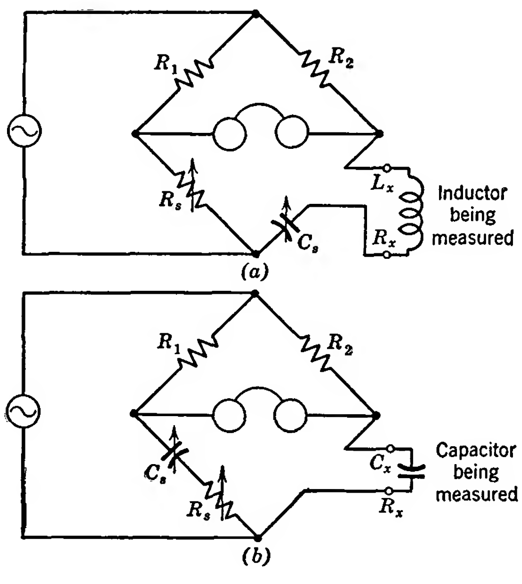


FIG. 21. An alternating-current bridge using standard capacitor C_s.

The bridge of Fig. 20(b) is adjusted until the inductive reactance $2\pi f L_s$ equals the capacitive reactance $1/(2\pi f C)$. For this condition,

$$C_x = \frac{1}{(2\pi f)^2 L_s}. \tag{43}$$

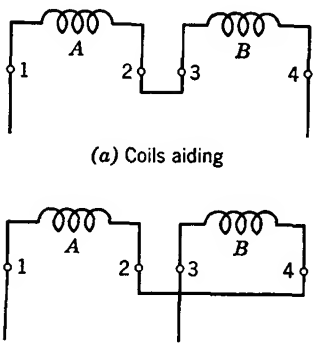
As will be noted, the frequency must be known.

Bridge with Standard Capacitor. For many purposes the standard capacitor of Fig. 21 has negligible effective resistance and no resistance corrections are necessary. For the circuit of Fig. 21(a), when resonance is obtained the value of the unknown inductor is

$$L_x = \frac{1}{(2\pi f)^2 C_s}. \tag{44}$$

If the circuit of Fig. 21(b) is used, then the capacitance of the unknown is given by equation 42.

Bridge Measurements of Mutual Inductance. The circuits of Figs. 20(a) and 21(a) are used for measuring mutual inductance. Suppose that the two coils of Fig. 22 are connected *aiding* so that the magnetic effects *add;* then, because of the mutual inductance between the coils, the back voltage between terminals 1 and 4 may be considered as composed of four components: (1) the back voltage caused by the self-inductance of coil A, (2) the voltage induced in coil A by the current in coil B and the mutual inductance between B and A, (3) the back voltage caused by the self-inductance of coil B, and (4) the voltage induced in coil B by the current in coil A and the mutual inductance between A and B. This total back voltage appears as an inductive effect at the terminals, and hence the equivalent self-inductance as measured by a bridge is



(a) Coils aiding

(b) Coils opposing

FIG. 22. Connections for determining mutual inductance.

$$L_{1-4} = L_1 + 2M + L_2\,. \tag{45}$$

But if the coils are connected opposing so that the magnetic effects *subtract*, then the induced voltages caused by the mutual inductances subtract from those caused by the self-inductances, and

$$L_{1-3} = L_1 - 2M + L_2\,. \tag{45a}$$

Subtracting equation 45*a* from equation 45 gives

$$L_{1-4} - L_{1-3} = 4M. \tag{46}$$

The mutual inductance between two coils is, accordingly, one-fourth the difference in the inductance measured with the two coils aiding and the two coils opposing.

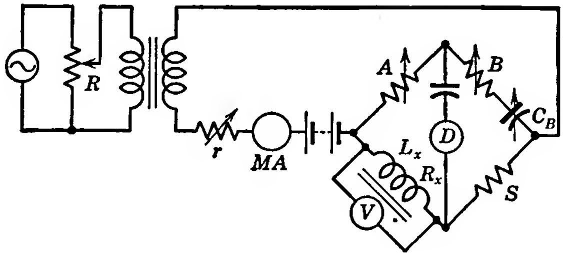


FIG. 23. Circuit for measuring the incremental inductance L_x and the effective resistance R_x of a coil with a core of iron or other ferromagnetic material.

Bridge Measurements of Incremental Inductance. The magnitudes of the incremental inductance and effective resistance of an inductor with a ferromagnetic core will vary with the magnitude of both the direct and the alternating currents through the coil (page 53). The bridge of Fig. 23 provides both direct current and alternating current. The direct current from the battery is regulated by resistance r and measured with the milliammeter. It is difficult to measure directly the magnitude of the alternating-current component, although it can be done. A high-impedance vacuum-tube voltmeter connected as indicated can be used to maintain constant the alternating voltage drop across the inductor. This is accomplished by varying the voltage divider R. Both the direct current and the alternating voltage must be kept constant as the bridge is balanced.

If the inductor has many turns (a filter choke for instance), then a low frequency of about 100 cycles must be used to avoid the effects of the distributed capacitance. For this reason, null detector D is often a tunable vacuum-tube amplifier and detector, or some similar device. It should be noted that resistor S must pass the direct-current component. At balance,[16]

$$L_x = ASC_B, \quad R_x = ABS\omega^2 C_B{}^2 \quad \text{and} \quad Q = \frac{1}{(B\omega C_B)} \text{ (approximately)}, \tag{47}$$

where all values are in henrys, farads, and ohms, and ω equals 2π times the frequency.

Thermocouples. Alternating-current and voltage measurements are made with thermocouples and associated galvanometers. Each thermocouple of Fig. 24 consists essentially of two wires (A, B) of dissimilar metal fastened together at the thermocouple junction. When the junction is heated with the alternating current to be measured, a direct voltage is produced across the galvanometer. This voltage is proportional to the heat dissipated, and hence to I^2R, where I is effective value of the alternating current flowing and R is the effective resistance of the thermocouple heater.

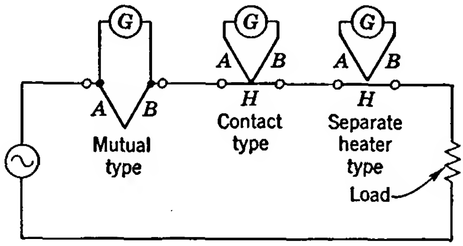


FIG. 24. Illustrating three types of thermocouples and how they would be connected to measure the current taken by a load. The dissimilar metals forming the thermocouple are A and B. The heater is marked H.

The **mutual type** is not very satisfactory. The alternating current to be measured divides, part passing through the galvanometer and part through the load. Since the impedance of these two parallel paths varies with frequency, the current division also varies, and a **shunting error** is introduced. Furthermore, a **reversal error** exists, causing a different thermocouple output for each direction of current flow.

In the **contact type** of thermocouple the shunting error is negligible, and the reversal error is very small and may usually be neglected. The calibration may readily be made with direct current and is independent of frequency except at very high frequencies where skin effect alters the effective resistance and the stray capacitance shunts current out of the heater.

The **separate-heater type** has no reversal error and is especially suitable for high-frequency measurements where the capacitance of the galvanometer to ground would be objectionable. Maximum sensitivity is obtained by binding the thermocouple junction and the heater together with a bead of heat-conducting but electrical insulating material.

At low frequencies, such as over the audio range, about the only effect of inserting a thermocouple in a circuit is to add resistance to the circuit. At radio frequencies, however, the equivalent circuit[19] for a separate-heater type thermocouple is as shown in Fig. 25. Because of the stray inductances and capacitances, a calibration made with direct current, or with low frequencies, does not hold at extremely high frequencies.

For measuring an alternating current, the thermocouple is inserted in the circuit as in Fig. 24. For measuring an alternating voltage, the thermocouple and a series resistance are connected as in Fig. 26. The voltage between the line wires is $E = I(R + r)$, where R is the protec-

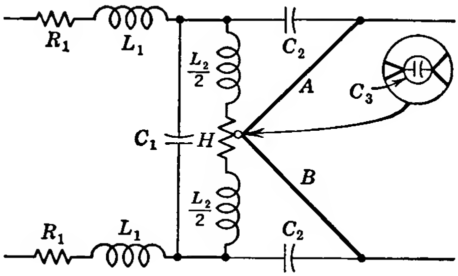


FIG. 25. The high-frequency equivalent circuit of a thermocouple of the separate-heater type. A and B are the thermocouple wires, and H is the heater. L_1 and L_2 are the inductances of the connecting leads and the heater. C_1 is the capacitance between connecting leads. C_2 is the capacitance between heater and thermocouple. C_3 is the capacitance of the insulating bead between the thermojunction and heater. Mutual inductance (not shown) exists between the heater and thermocouple. (Adapted from Reference 19.)

tive or current-limiting resistor, r is the resistance of the thermocouple heater (this should be accurately measured) and I is the current obtained from the calibration curve. These arrangements take power from the source and accordingly alter circuit conditions. There are instances when thermocouples cannot be used and when vacuum-tube measuring devices must be employed.

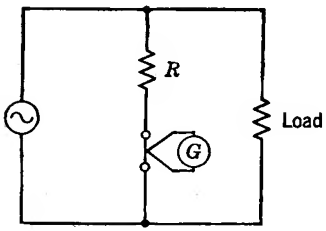


FIG. 26. Methods of measuring voltage with a thermocouple.

Shielding. If instruments, or circuits such as bridges, are operated in weak constant electric or magnetic fields, difficulties are seldom experienced. If these fields are strong, however, or if they are alternating rather than constant, then serious troubles are experienced.

A circuit or an instrument can be shielded from the effect of a constant (direct) electric field by enclosing the object to be shielded in a metal box.[20] Since the metal is electrically conducting, the lines of electric force do not penetrate to the interior but terminate on surface charges. An instrument or circuit can be shielded from the effects of a constant magnetic field by entirely enclosing the device in a thick case of good

magnetic material such as iron. The magnetic lines of force follow through the metal of the shield and penetrate less within.

Although shielding against stray alternating-current fields is in reality a process of shielding against electromagnetic waves, it is convenient to consider the separate components.[21] There are, therefore, two problems: *first,* shielding against stray alternating *magnetic* fields; and *second,* shielding against stray alternating *electric* fields.

Shielding against stray alternating magnetic fields[22] is usually limited to wound apparatus such as transformers. Often the coils are wound on a closed core of some good magnetic material such as iron or Permalloy; then the flux will be confined almost entirely to the magnetic core, and but little coupling to other circuits will be experienced. If such closed cores are not sufficient or are not used, there are two other methods of shielding.

The *first* of these is to enclose the device in a case of magnetic material as previously explained. This shield will prevent the magnetic lines of force from leaving the vicinity of the coil being shielded, and from coupling with another unit. The *second* method of magnetic shielding is to enclose the device in a box of non-magnetic material, such as copper, having high electrical conductivity. Then the alternating magnetic field produces eddy currents in the copper, and the magnetic field produced by these currents reacts with, and largely neutralizes, the stray field from the device tending to cause interference. Such shields cause losses and change the electrical characteristics of the apparatus shielded.

A metal box will shield against stray alternating electric fields. If it is desired to shield the individual pieces of equipment in a network (such as an impedance bridge) from the mutual effects of the other pieces, the problem becomes more difficult. Shielding against an electric field does not consist of *preventing* mutual coupling, but of *controlling* the coupling. Two problems are involved, depending on whether the units are in series or in parallel.

To illustrate the action of a shield, consider the impedance unit of Fig. 27(a). When an alternating electromotive force is applied between A–B, in addition to the current from A to B through the impedance, current flows from A to B through the stray capacitances and the grounded path as shown by the arrow. These stray capacitances and the resulting currents vary for every different position of the impedance with respect to ground, and with respect to surrounding objects such as an observer. If, however, a conducting shield is placed completely around the impedance as in Fig. 27(b), the stray paths will be made definite, and the impedance between the

points A–B will be independent of the location of the unit. Stray capacitance will also exist as shown between the shield and ground, but, since all parts of the shield are at the same potential, no current will flow.

Suppose that two elements are connected in series and that it is desired to shield these from each other, and also so that their impedances are independent of their position with respect to ground and surrounding objects. It might at first appear that connecting them as in Fig. 28(a) would be satisfactory. It is not, however;

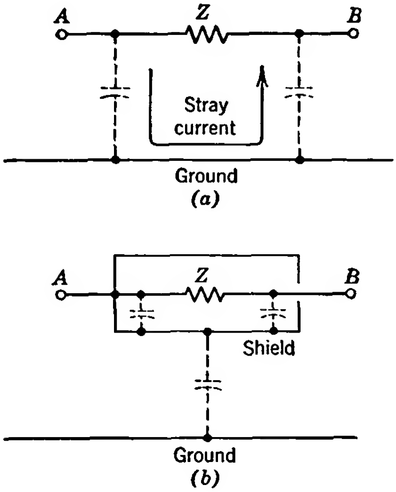


FIG. 27. Method of shielding a single element Z by the use of a conducting metallic shield. The impedance between A–B is independent of position with respect to ground. Element Z is also shielded from high-frequency electromagnetic fields.

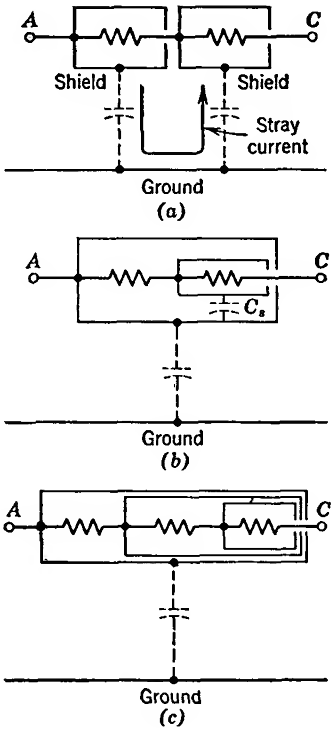


FIG. 28. Incorrect method (a) of shielding elements from each other, and from the effects of ground and surrounding objects. Correct methods are shown in (b) and (c).

the two shields are at different potentials, and current will flow as indicated by the arrow. The impedance between A–C will therefore be variable. If, however, the shield is extended as shown in Fig. 28(b), then the stray capacitances between the two shields, indicated by the dotted condenser C_s, will be constant, and the impedance between A–C will be independent of location. This arrangement can be extended to

include any number of elements, three elements being shielded in Fig. 28(*c*).

Attention is called to the fact that, although shielding tends to make a unit independent of position, it increases the impedance variations with frequency. Thus, if the unit of Fig. 27(*a*) is a resistor, the actual circuit between the terminals *A–B* consists of resistance and capacitance in parallel. This assumes, of course, that the inductance of the resistor is negligible. When the shield is placed in position, the capacitance is increased. Even at voice frequencies the circuit *A–B* would not be strictly pure resistance, and the higher the frequency, the greater is the deviation from this ideal. Shielding, therefore, *does not prevent stray fields* and capacitive coupling, but *controls* these, making them independent of position.

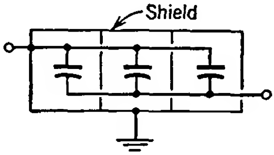


Fig. 29. Methods of shielding parallel elements.

Units connected in parallel should be individually shielded, and all the shields connected to some common point which is then usually grounded. This method is illustrated by Fig. 29.

Distortion. This is defined[1] as a "change in wave form." Three types of distortion are as follows:

Frequency Distortion. This is defined[1] as "that form of distortion in which the change is in the relative magnitudes of the different frequency components of a wave, provided that the change is not caused by non-linear distortion." Thus, if the transmitting efficiency of a circuit or piece of equipment is different at various frequencies, frequency distortion results. This is sometimes called amplitude distortion.

Non-Linear Distortion. This is defined[1] as "that form of distortion which occurs when the ratio of voltage to current, using root-mean-square values (or analogous quantities in other fields), is a function of the magnitude of either." Non-linear distortion causes harmonics to be created (see page 558). If a pure sine-wave voltage is impressed on a circuit (such as an iron-cored transformer) having a **non-linear impedance** (ratio of voltage to current as just defined), the current that flows will not be a pure sine wave, but will contain harmonics and will be irregular in shape. Non-linear circuits have many useful applications in communication, such as modulation (page 420), demodulation (page 421), and the generation of carrier frequencies (page 425).

Delay Distortion. This is defined[1] as "that form of distortion which occurs when the phase angle of the transfer impedance (page 423) with respect to two chosen pairs of terminals is not linear with frequency

within a desired range, thus making the time of transmission or delay vary with frequency in that range." Delay distortion is the type of distortion that occurs when the velocity of propagation or speed of transmission of a wave through a circuit varies with the frequency of the wave. Considerable delay distortion is tolerable in the transmission of speech and music but is very objectionable in picture and television transmission.

The Decibel. The **bel** is defined[1] as "the fundamental division of a logarithmic scale for expressing the ratio of two amounts of power, the number of bels denoting such a ratio being the logarithm to the base 10 of this ratio." If P_1 and P_2 designate two amounts of power, and if N is the number of bels denoting their ratio,

$$N = \log_{10} \frac{P_1}{P_2}. \tag{48}$$

The **decibel** is defined[1] as "one-tenth of a bel, the number of decibels denoting the ratio of two amounts of power being 10 times the logarithm to the base 10 of this ratio." The abbreviation **db** is used extensively. If P_1 and P_2 designate two amounts of power, and if n is the number of decibels denoting their ratio,

$$n = 10 \log_{10} \frac{P_1}{P_2}, \quad \text{and} \quad \frac{P_1}{P_2} = 10^{0.1 \times n}. \tag{49}$$

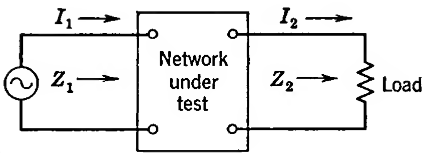


FIG. 30. Circuit for illustrating the meaning of the term decibel.

Power is seldom measured directly in communication circuits. It is usually determined indirectly from current measurements made with thermocouples, or from voltage measurements made with vacuum-tube voltmeters. Because of these indirect measurements of power, confusion has resulted in the use of the decibel, and the reasons will now be explained.

It is desired to determine the *power* loss in decibels of the network of Fig. 30. This is to be done from the ratio of the input current I_1 and the output current I_2. Since power is I^2R, equation 49 can be written

$$n = 10 \log_{10} \frac{I_1^2 R_1}{I_2^2 R_2} = 20 \log_{10} \frac{I_1}{I_2} + 10 \log_{10} \frac{R_1}{R_2}. \tag{50}$$

If $R_1/R_2 = 1.0$, then the last term equals zero, and *under these special*

conditions, in which the resistance components of the input impedance equals the resistance component of the load impedance, the power ratio and the loss in decibels is

$$n = 20 \log_{10} \frac{I_1}{I_2}, \quad \text{and} \quad \frac{I_1}{I_2} = 10^{0.05 \times n}. \tag{51}$$

Similarly, if the *power loss* is to be determined from voltage readings, where E_1 is the voltage impressed across the network and E_2 is the voltage across the load, since power $= EI \cos \Theta$, equation 49 can be written

$$n = 10 \log_{10} \frac{E_1 I_1 \cos \Theta_1}{E_2 I_2 \cos \Theta_2} = 10 \log_{10} \frac{E_1{}^2 Z_2 \cos \Theta_1}{E_2{}^2 Z_1 \cos \Theta_2} =$$

$$20 \log_{10} \frac{E_1}{E_2} + 10 \log_{10} \frac{Z_2}{Z_1} + 10 \log_{10} \frac{\cos \Theta_1}{\cos \Theta_2}. \tag{52}$$

If it happens that Z_1 equals Z_2 in both magnitude and phase, then the last two terms become zero, and *under these special conditions* the power ratio and the loss in decibels is

$$n = 20 \log_{10} \frac{E_1}{E_2}, \quad \text{and} \quad \frac{E_1}{E_2} = 10^{0.05 \times n}. \tag{53}$$

The confusion mentioned previously resulted from using equations 51 and 53 to determine *power ratios* in decibels in circuits that were not *matched* and did not meet the special conditions explained. Most telephone lines and apparatus *are* matched for maximum power transfer, and thus the impedance relations are such that equations 51 and 52 can be used without error. However, many circuits and much equipment are not matched for maximum power transfer, and hence these equations do not apply.

It is common practice to use equations 51 and 53 in circuits in which the impedances are *not* matched, but when so used these equations *do not give power ratios* in decibels but instead express *current or voltage ratios* in decibels. This is in accordance with the Standards,[1] but the usage should be accompanied by a statement of the specific application.

Gains in decibels as well as losses are expressed by equations 49, 51, and 53. In all instances the larger number is divided by the smaller, and the terminology "gain" or "loss" is used.

Equations 49, 51, and 53 can be plotted as in Fig. 31. With these

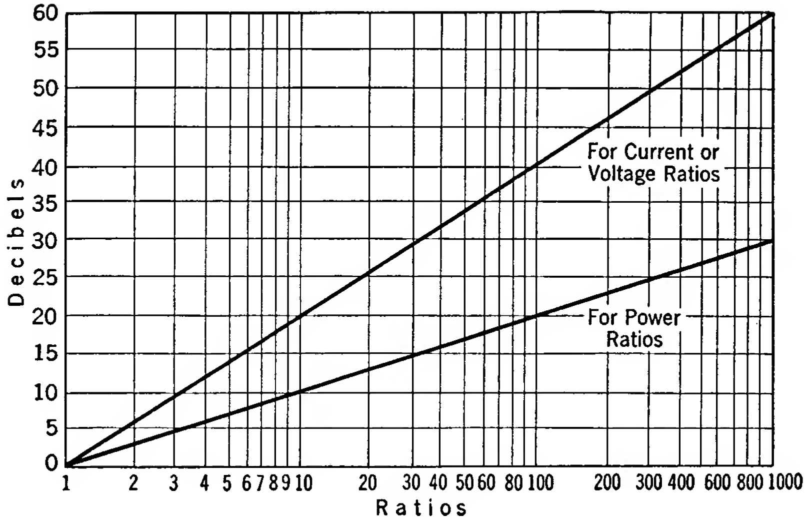


FIG. 31. Curves for gain or loss in decibels, from current, voltage, or power ratios.

curves, the gain or loss in decibels can be found. Also, gains or losses can be determined from Table V.

TABLE V

RELATIONS BETWEEN DECIBELS AND FRACTIONAL LOSS OR GAIN IN POWER

Decibels	Loss-Power Ratio	Gain-Power Ratio
1	0.794	1.26
2	0.631	1.58
3	0.501	2.00
4	0.398	2.51
5	0.316	3.16
6	0.251	3.98
7	0.200	5.01
8	0.158	6.31
9	0.126	7.94
10	0.100	10.00
20	0.010	100.00
30	0.001	1,000.00
40	0.0001	10,000.00
50	0.00001	100,000.00

Transmission losses are sometimes expressed in nepers. From equations 51 and 53, the loss in decibels is

$$n = 20 \log_{10} \frac{I_1}{I_2}, \quad \text{or} \quad n = 20 \log_{10} \frac{E_1}{E_2}.$$

If the natural or Napierian base for the logarithms is used,

$$n' = \log_\epsilon \frac{I_1}{I_2}, \quad \text{or} \quad n' = \log_\epsilon \frac{E_1}{E_2}, \tag{54}$$

where n' is the loss (or gain) in nepers. The relation between the natural and common logarithms is $\log_\epsilon x = 2.3026 \log_{10} x$. Therefore, from equation 54, $n' = 2.3026 \log_{10} I_1/I_2$. From equation 51, $n/20 = \log_{10} I_1/I_2$, and

$$\left.\begin{aligned} &\text{Loss in nepers} &&= 0.11513 \text{ times loss in decibels.} \\ &\text{Loss in decibels} &&= 8.686 \text{ times loss in nepers.} \end{aligned}\right\} \tag{55}$$

Power Level and Volume. If a given reference value is chosen, the **power level,** or power being transmitted past any point in a system,[1] may be expressed as so many decibels above or below the selected reference. For several years the reference point or **zero level** chosen was approximately 0.006 watt, or 6 milliwatts. On this basis, if a circuit were delivering 60 milliwatts, it was operating at a "plus 10 db power level." Or, if operating at a "minus 10 db level," the power at that point would be 0.6 milliwatt. A zero level of 0.001 watt is common.

It is also common practice to select some strategic point in the circuit as a reference point and express the power level in decibels in other parts with respect to this arbitrary point. Thus, care must be exercised in interpreting data because of the different zero levels used. Because of this, there is a growing tendency to use the designation **dbm** when the zero level is 1.0 milliwatt (or 0.001 watt).

Power levels are used in circuit tests where steady-state sine-wave currents are employed. Such factors as the ballistic characteristics of the measuring instrument do not, therefore, affect the measured values. For measuring program levels on circuits transmitting speech and music, where the currents are of transient nature, the characteristics of the instrument do affect the measured value.

The **volume** at any point in a telephone circuit is a measure of the power of a voice-frequency wave at that point.[1] Volume is expressed in decibels with respect to some arbitrary reference standard. A special **volume-level indicator** or VI has been developed[2,3] for measurements in program circuits. This reference specifies (*a*) the *characteristics* and *method of use* of the volume-indicator instrument, and (*b*) that the steady-state reference power is 1.0 milliwatt in a circuit of 600-ohms characteristic impedance (see page 202). The term **volume units** (abbreviated vu) is used on the scale of the volume indicator, one volume unit denoting that the program volume level is one decibel above zero reference volume.

REFERENCES

1. *Definitions of Electrical Terms.* 1941. (A.S.A. C42) A.I.E.E.
2. Rader, L. T., and E. C. Litscher. Some Aspects of Inductance when Iron is Present. *Elec. Eng.,* March, 1944, Vol. 63, No. 3.
3. Terman, F. E. *Radio Engineers' Handbook,* First Edition. McGraw-Hill Book Co.
4. Elmen, G. W. Magnetic Alloys of Iron, Nickel, and Cobalt. *Elec. Eng.,* December, 1935, Vol. 54, No. 12.
5. Legg, V. E. Survey of Magnetic Materials and Applications in the Telephone System. *Bell System Tech. J.,* July, 1939, Vol. 18, No. 3.
6. Supermalloy. *Bell Lab. Record,* March, 1948, Vol. 26, No. 3.
7. Martowicz, C. T. Powdered Iron Cores. *Electronic Inds.,* June, 1945, Vol. 4, No. 6.
8. Shea, H. G. Magnetic Powders. *Electronic Inds.,* August, 1945, Vol. 4, No. 8.
9. Beller, H., and G. O. Altmann. Radio Frequency Cores of High Permeability. *Electronic Inds.,* November, 1945, Vol. 4, No. 11.
10. Williams, C. S. Permanent Magnet Materials. *Elec. Eng.,* January, 1936, Vol. 55, No. 1.
11. An Important New Magnetic Alloy. *Bell Lab. Record,* September, 1940, Vol. 19, No. 1.
12. Elmen, G. W. Magnetic Properties of Perminvar. *Bell System Tech. J.,* January, 1929, Vol. 8, No. 1.
13. *General Radio Experimenter,* June, 1939, General Radio Co.
14. Everitt, W. L. *Communication Engineering.* McGraw-Hill Book Co.
15. Tuttle, W. N. The Series and Parallel Components of Impedance. *General Radio Experimenter,* January, 1946, General Radio Co.
16. Hague, B. *Alternating Current Bridge Measurements.* Pitman & Son.
17. Sinclair, D. B. The Twin-T—A New Type of Null Instrument for Measuring Impedance at Frequencies up to 30 Megacycles. *Proc. I.R.E.,* July, 1940, Vol. 28, No. 7.
18. Sinclair, D. B. A Radio-Frequency Bridge for Impedance Measurements from 400 Kilocycles to 60 Megacycles. *Proc. I.R.E.,* November, 1940, Vol. 28, No. 11.
19. *Catalog K,* General Radio Co.
20. Curtis, H. L. Shielding and Guarding Electrical Apparatus used in Measurements. *Trans. A.I.E.E.,* October, 1929, Vol. 48, No. 4.
21. Ferguson, J. G. Shielding in High-Frequency Measurements. *Trans. A.I.E.E.,* October, 1929, Vol. 48, No. 4.
22. Gustafson, W. G. Magnetic Shielding of Transformers at Audio Frequencies. *Bell System Tech. J.,* July, 1938, Vol. 17, No. 3.
23. Affel, H. A., H. A. Chinn, and R. M. Morris. A New "VI" and Reference Level. *Electronics,* February, 1939.

REVIEW QUESTIONS

1. Define effective resistance. In alternating-current work, why should the old "Ohm's law" definition of resistance be avoided?
2. Explain why the effective resistance of a circuit may differ from the direct-current resistance.

3. A coil has a laminated silicon-steel core. Would you expect the permeability of the core to be exactly the same at 50 cycles as at 10,000 cycles, all other factors being comparable? Explain why.
4. Discuss the nature of the inductance of a coil with a ferromagnetic core.
5. What is the effect of a small air gap on the incremental inductance of a coil with a ferromagnetic core?
6. When phase relations are of importance, why should $-j$ be written before equation 8?
7. On page 53 a somewhat detailed explanation was given of the inductive effects in coils with ferromagnetic cores. What parts of this discussion apply to the mutual inductance of coils with such cores?
8. Would you attempt to use a 10,000-ohm wire-wound resistor at a frequency of 1.0 megacycle? Why? If your answer is negative, what type would you use?
9. In telephone equipment non-inductive resistors are often made by winding the resistance wire on a thin card. Fully explain why this reduces the inductance. What are other advantages of a resistor of this type?
10. Discuss the types of cores used in communication inductors and transformers.
11. Discuss the types of dielectrics used in communication capacitors.
12. What is meant by the "Q" of an inductor and "D" of a capacitor? Are these two terms related?
13. Briefly discuss the magnetic materials used in communication.
14. Briefly discuss the dielectric materials used in communication.
15. Enumerate and briefly discuss the various frequency bands used in communication.
16. On pages 64 and 65, *phase* resonances were discussed, the term *phase* being used to prevent confusion with other types of resonance. What other possible types of steady-state electrical resonances are there? Why is the term antiresonance sometimes used?
17. In solving an inductively coupled circuit as explained on page 67, the effect of the secondary reflected into the primary is considered. Explain why it is not necessary to consider the effect of the primary reflected into the secondary.
18. What are the conditions for maximum power transfer in an alternating-current circuit? Why is maximum power transfer and not efficiency often of greatest importance in communication?
19. Why are impedance transformations often made in communication? Name several methods of transforming impedances.
20. In deriving the equations for an alternating-current bridge, why are the in-phase and out-of-phase terms separated?
21. A device called an Inductometer or Variometer is constructed on the principle of the circuit of Fig. 18(a). It is used to furnish the variable inductance for bridge standards. Explain how such a device should be constructed to give variable inductance and constant effective resistance. Could the device have zero inductance?
22. In making bridge measurements of incremental inductance why should a thermocouple not be placed directly in series with the coil under test?
23. In discussing thermocouples, it was stated that the separate-heater type was used for high-frequency measurements. Explain why on the basis of the theory on page 139.
24. Explain why a copper or aluminum shield is effective against a high-frequency

magnetic field. Will such a shield affect the resistance and inductance of an air-cored coil? If so, how?

25. Under what conditions can power ratios in decibels be found from current and voltage ratios? Is the decibel used for purposes other than measuring power ratios? Name several.

PROBLEMS

1. An air-core coil has a self-inductance of 0.056 henry, and a resistance of 9.2 ohms. Calculate the Q at 100, 1000, 10,000, 1,000,000, and 10,000,000 cycles. Plot as a curve. Would measured values agree with calculated values? Why?
2. If the coil just considered is connected in series with a 0.5-microfarad capacitor, what will be the resonant frequency? What will be the impedance at resonance, and at 10 per cent above and below resonance?
3. Repeat Problem 2 for the inductor and capacitor in parallel. Use both equations 15 and 16, and compare.
4. An oscillator has an open-circuit voltage of 39 volts at 1000 cycles and an internal impedance of 0.003 henry inductance and 447 ohms resistance. It is connected to the primary of a coil having an inductance of 0.052 henry and a resistance of 56 ohms. The mutual inductance between the primary and the secondary is 0.0536 henry. The secondary has an inductance of 0.061 henry, a resistance of 61 ohms, and is connected to a 500-ohm load resistor. Calculate the current that will flow through this resistor.
5. If the combination of Problem 2 is connected across 50 volts at the resonant frequency, what will be the line current, and the voltage across the inductor and across the capacitor?
6. If the combination of Problem 3 is connected in parallel and across 50 volts at the resonant frequency, what will be the line current, and the current through the inductor and capacitor?
7. Prove mathematically that the statement in italics preceding equation 24 is correct.
8. Prove mathematically that the statement in italics following equation 29 is correct.
9. Repeat the problems starting on page 72 for increasing and decreasing impedances, but with frequencies of 2.0 megacycles instead of 5.0 megacycles, and at 7.5 instead of 10.0 megacycles.
10. The heater of a thermocouple has a constant resistance of 610 ohms. For a current of 0.002 ampere through the heater, the deflection of the associated microammeter is 10 microamperes. The thermocouple heater is placed in series with 9390 ohms, and placed across an unknown voltage. The microammeter reads 100 microamperes. What is the value of the unknown voltage in volts?

CHAPTER 4

ELECTROACOUSTIC DEVICES

Introduction. If the spoken word or a program of music is to be transmitted electrically, some electroacoustic device must either *convert* the sound waves into electric energy or permit the sound waves to *control* the electric power supplied by a source such as a battery. Such an electroacoustic device is defined[1] as a **telephone transmitter,** "a device whereby sound waves produce substantially equivalent electric waves." The familiar word **microphone** is[1] "a term frequently used as a synonym for telephone transmitter, particularly in radio and sound picture fields."

A telephone transmitter (or microphone) is a special form of **electroacoustic transducer,** defined[1] as "a transducer by which power may flow from an electric system to an acoustic system or vice versa." The term **transducer** is defined[1] as "a device by means of which energy may flow from one or more transmission systems to one or more other transmission systems." There are two types of transducers, a **passive transducer,** containing no source of power, and an **active transducer,** containing a source, or sources, of power.[1] Accordingly, telephone transmitters (or microphones) that contain no source of power but which convert acoustic energy into electric energy are passive or generator-type transmitters and sometimes are called **passive electroacoustic transducers.**[2] Similarly, telephone transmitters that use sound waves to control the flow of power from a battery are active or modifier-type transmitters and are sometimes called **active electroacoustic transducers.**[3]

Passive or Generator-Type Telephone Transmitters. In these transmitters (or microphones) all the electric power output must come from the acoustic power in the sound waves. The passive transmitters are, therefore, essentially electric generators with internal impedance and an open-circuit output voltage that is a replica of the sound waves. Passive or generator-type transmitters can be grouped in two classes: the **magnetic microphone,** defined[1] as "a microphone the electric output of which is generated by the relative motion of a magnetic field and a coil or conductor located within the magnetic field," and the

crystal microphone, which is a microphone that operates by virtue of the piezoelectric effect. These devices will be considered in detail later in this chapter.

The electric power output of a passive or generator-type transmitter is very low, because the electric power output comes entirely from the acoustic power of about 10 microwatts actuating the transmitter, and because the efficiency of transforming from acoustic to electric power is a few per cent.

Active or Modifier-Type Telephone Transmitters. In active telephone transmitters the sound waves *control* an external source of power. For example, in the telephone set, the transmitter modifies the current flowing from a battery. Such transmitters are also amplifiers. The common telephone transmitter may deliver an electric power output 1000 times the magnitude of the acoustic power from the voice of the speaker.[4] The power output of most modifier transmitters is accordingly many times greater than that of the generator type.

Many types of modifier transmitters have been developed,[5] including liquid transmitters,[5] flame transmitters,[5] glow transmitters,[6] the thermophone,[5] carbon-granule transmitters, and condenser transmitters. The last two are of much practical interest and will be discussed in detail.

The Single-Button Carbon-Granule Telephone Transmitter. Many types of telephone transmitters have been developed,[7] one being shown in Fig. 1. This consists of a circular brass cup partly filled with carefully selected carbon granules. The inside edge of the cup is insulated with a paper ring. In the bottom is placed a highly polished, circular, carbon electrode which is fastened to the cup and serves for a contact with the carbon granules. The opening or top of the cup is partly filled with another circular electrode fastened to a metal plunger which in turn is attached to the diaphragm.

The path of the current is through the insulated terminal fastened to the steel bridge, through the damping spring, through the front movable disk, through the carbon granules, and thence out through the fixed carbon disk to a second insulated binding post on the steel bridge. The damping spring is provided to reduce resonance effects in the mechanical portions. The resistance of the transmitter at a current of about 0.1 ampere is about 50 ohms when it is not excited by sound waves.

The Functioning of the Carbon Granules. It is thought that microphonic action is the result of variations of contact area with deformation of the contact material.[8, 9] The carbon granules act like elastic spheres which flatten out as they are pressed together.

The resistance decreases with an increase of the area of the contact. It is also thought that mechanical forces on the granules establish new contacts as well as change the areas of those already formed.

A special type of carbon is used for the granules, and these must be specially selected as to size and quality. They must be free from corrosion and of high resistance, and their contact resistance must be sensitive to pressure changes. If the granules are not of good quality

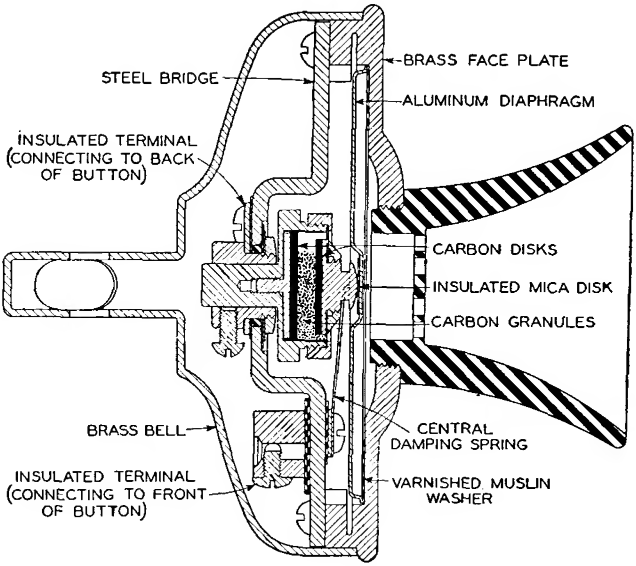


FIG. 1. Cross section of a Western Electric carbon-granule transmitter. Although this type will gradually be replaced by those of later design, many transmitters of this general type will remain in useful service for years.

or if they are not carefully selected, they readily become **packed** and the transmitter becomes insensitive. Packing may also be caused by the entrance of moisture, by overheating, or by other misuses. *Slightly* jarring the transmitter will usually remedy this condition.

If the transmitter of Fig. 1 is held at an angle, the granules tend to fall away from the electrodes, and this may completely open the circuit. It is important that the transmitter be held close to the speaker's lips when it is being used. One advantage of the modern handset telephone is that, if the receiver is in position for hearing, the transmitter is in the correct position for talking.

When the current through the carbon transmitter is increased beyond a certain value, the points where the currents enter and leave the

granules will become hot and arcing will take place. A hissing or frying sound[10] will be heard, and the so-called frying or burning point of the transmitter has then been reached. The normal current taken by telephone transmitters is about 0.1 ampere. In certain water-cooled carbon-granule transmitters of historical interest 15 amperes have been carried continuously.[7]

Transmitters for Telephone Handsets. The telephone **handset** is defined[1] as a "combination of a telephone transmitter and a telephone

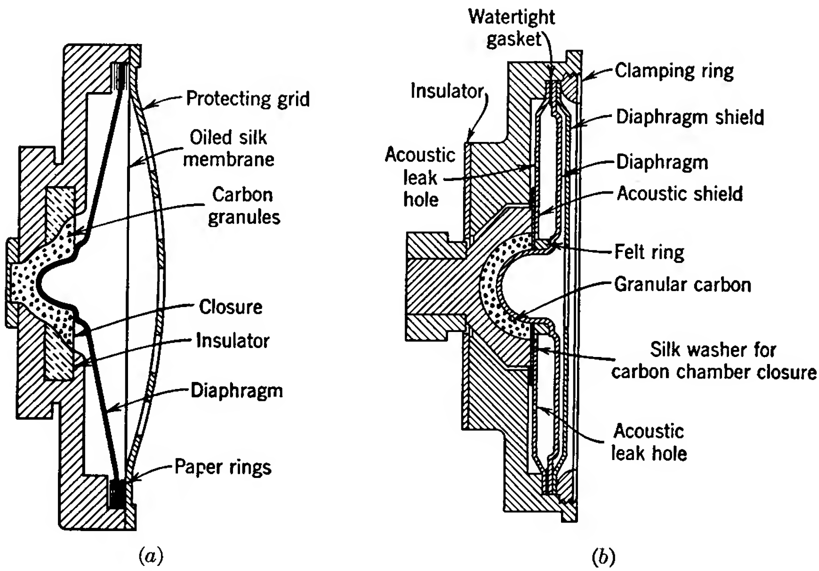


Fig. 2. Construction of typical telephone handset transmitters. (*a*. Courtesy Western Electric Co.; *b*. courtesy Kellogg Switchboard and Supply Co.)

receiver mounted on a handle." The transmitter electrodes and the carbon-granule cup must be constructed so that the granules cannot fall away from the electrodes and open the circuit for any position in which the handset is held; that is, the transmitter must be *non-positional.*

Typical methods of construction are shown in Fig. 2. In each of these the diaphragm is formed and placed so that it acts as the front electrode. The other electrode is at the back of the carbon-granule cup or container. The cups are not filled *entirely* with granules because space must be left for expansion of the granules when the temperature rises. The diaphragm of Fig. 2(*a*) is cone shaped and ribbed so that it will be stiff and will move in and out somewhat like a

piston. The diaphragm of Fig. 2(*b*) is damped acoustically so that it does not vibrate excessively at certain resonant frequencies.

A second type of handset transmitter is shown in Fig. 3(*a*). The diaphragm consists of two thin aluminum-alloy cones. The two

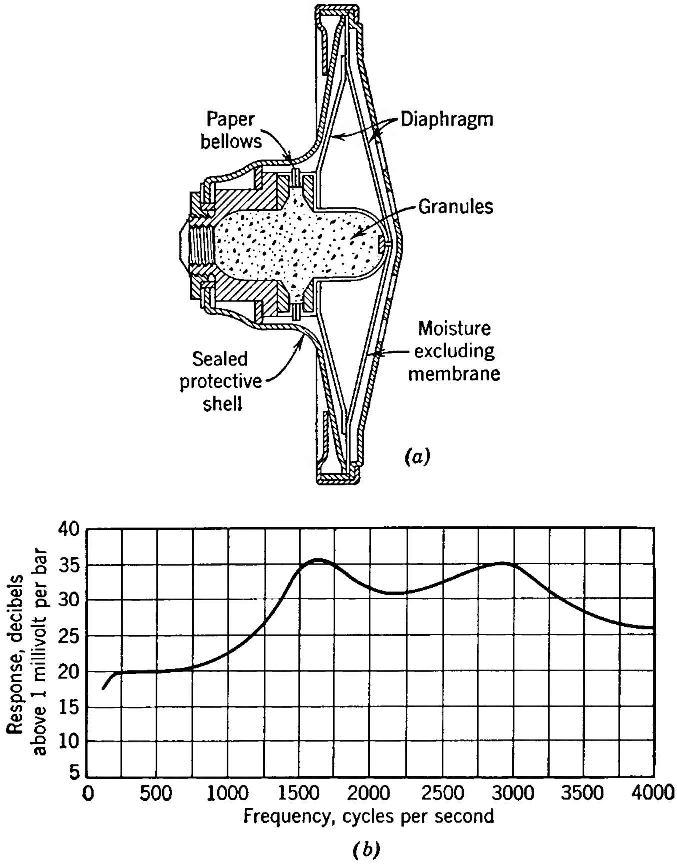


Fig. 3. Construction (*a*) and frequency response (*b*) of a typical telephone handset transmitter. For the meaning of the word bar, see Fig. 9. (Courtesy Automatic Electric Co.)

electrodes are separated by paper bellows. The frequency-response curve is shown in Fig. 3(*b*).

The telephone transmitters of Figs. 2 and 3 are known as "capsule types" because they are made as a unit and cannot be adjusted in the field. The characteristics of the handset telephone transmitters are superior to those of the transmitter of Fig. 1.

Telephone Transmitter Operation in Resistive Circuits. In this section the theory of a single-button, carbon-granule transmitter in a circuit containing resistance only will be discussed. It is assumed that the transmitter diaphragm is excited by a pure tone; also, that the diaphragm of the transmitter and the resistance of the carbon granules vary harmonically.

The total resistance at any instant of the circuit of Fig. 4 is

$$R = (R_c + R_t) - r \sin \omega t. \tag{1}$$

In this expression, R_t is the steady-state transmitter resistance, R_c is the resistance of the remaining portion of the circuit, and r is the maximum variation in transmitter resistance from R_t. When the

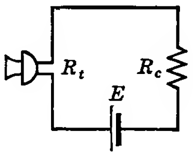


FIG. 4. Telephone transmitter in resistive circuit.

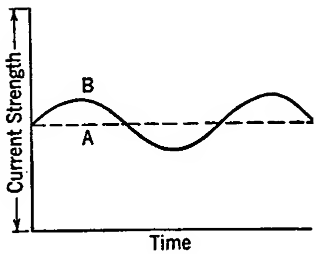


FIG. 5. Line A represents direct-current component or current flow when diaphragm is at rest. Line B represents current flowing when diaphragm is excited by a pure tone.

sound-wave pressure is maximum, the diaphragm will be pressed in and the circuit resistance will be $(R_c + R_t) - r$. When the sound-wave pressure is least, the transmitter diaphragm will be bent out and the circuit resistance will be increased to the value $(R_c + R_t) + r$.

The current through the transmitter at any instant will be similar to that shown by Fig. 5. It consists of direct and alternating components and is given by the relation

$$I_t = \frac{E}{(R_c + R_t) - r \sin \omega t} \quad \text{or} \quad \frac{E}{\left(R_c + R_t\right)\left(1 - \dfrac{r}{(R_c + R_t)} \sin \omega t\right)}. \tag{2}$$

This equation is in the form of the series

$$\frac{1}{1 - x} = 1 + x + x^2 + x^3 \cdots + x^n. \tag{3}$$

Equation 2 may be expanded according to this series, becoming

$$I_t = \frac{E}{(R_c + R_t)}\left(1 + \frac{r}{(R_c + R_t)}\sin \omega t + \frac{r^2}{(R_c + R_t)^2}\sin^2 \omega t + \cdots\right). \quad (4)$$

This equation indicates that, in addition to current variations having the same frequency as the impinging sound waves $\left(\frac{r}{(R_c + R_t)}\sin \omega t\right)$, harmonics, such as obtained when $\left(\frac{r^2}{(R_c + R_t)^2}\sin^2 \omega t\right)$ is expanded, are produced. These harmonics were not present in the original variations assumed, and thus distortion has been produced, although it is usually not serious. In fact, all terms higher than the first may usually be neglected, and the total instantaneous transmitter current can be written

$$I_t = \frac{E}{(R_c + R_t)}\left(1 + \frac{r}{(R_c + R)}\sin \omega t\right), \quad (5)$$

or

$$I_t = I_o\left(1 + \frac{r}{(R_c + R_t)}\sin \omega t\right), \quad (6)$$

where I_o is the normal or steady-state current flowing when the diaphragm is at rest. This equation becomes

$$I_t = I_o + \left(\frac{I_o r}{(R_c + R_t)}\sin \omega t\right). \quad (7)$$

From a telephone standpoint, interest is in the largest value that the alternating component may have, that is, in the portion $\frac{I_o r}{(R_c + R_t)}$. Since $I_o = \frac{E}{(R_c + R_t)}$, the maximum value of the alternating component becomes, from equation 7,

$$I_{\max} = \frac{Er}{(R_c + R_t)^2}. \quad (8)$$

To illustrate the application of this equation, suppose that a number of transmitters of different normal resistance R_t are available for connection to a circuit of resistance R_c, and also assume that for each one the ratio $r/R_t = p$ is a constant value. That is, the *percentage change of resistance* r when excited by the same source is the same. Then, equa-

tion 8 becomes

$$I_{\max} = \frac{EpR_t}{(R_c + R_t)^2}. \tag{9}$$

Now let it be assumed that all the factors in this equation are constant except R_t and that the expression is differentiated with respect to R_t. Then,

$$\frac{dI_{\max}}{dR_t} = \frac{Ep(R_c + R_t)^2 - EpR_t(2R_c + 2R_t)}{(R_c + R_t)^4}. \tag{10}$$

The greatest value of the alternating component $I_{\max}$ will be obtained if equation 10 is equated to zero. Thus,

$$Ep(R_c + R_t)^2 - EpR_t(2R_c + 2R_t) = 0,$$

and

$$R_t = R_c. \tag{11}$$

For the maximum useful alternating (speech) component of current in a circuit the steady-state resistance of the transmitter R_t should equal the resistance of the external circuit. If these relations are obtained, the useful output can be increased by substituting a transmitter having a greater change in resistance r.

Distortion Produced by Telephone Transmitters. The single-button carbon-granule telephone transmitters do not have the high quality of radio microphones, but they do have what is much more important—greater power outputs. In a telephone system much distortion can be tolerated, provided that the intelligibility is satisfactory and too much unnaturalness does not result.

It was shown by equation 4 that distortion resulted from normal transmitter action. For transmitters of given resistance R_t in circuits of resistance R_c, the least sensitive transmitters will produce the least distortion since they have the smallest total resistance change r. It follows that a given transmitter in different circuits will produce the least distortion in high-resistance circuits. In such circuits a given transmitter will, of course, produce less useful output.

Much distortion is caused by the diaphragm which is mechanically resonant at certain frequencies. At these frequencies the diaphragm motion and the electric output will be greater than for sounds of the same intensity but different frequency. In modern transmitters the diaphragms are ribbed or otherwise stiffened so that flexure is reduced and piston motion is approximated. The resonant cavities caused by the air spaces around the diaphragm and by the mouthpiece also cause distortion unless these cavities are designed properly.

A small amount of distortion is caused by the unequal travel of the diaphragm. When moving inward it meets an increasing opposition as it tends to compress the carbon granules, but this opposition does not occur when the diaphragm moves outward. This is an important factor leading to the use of the double-button carbon-granule microphone to be considered later.

The Magnetic "Sound-Powered" Telephone Transmitter. The sound-powered transmitter is of the generator type. Acoustic power from the sound waves is converted into electric power; hence the wording "sound-powered." As explained on page 94, the electric power output is very low, and such transmitters are used only in special telephone systems where the talking distances are small and the transmission losses are low.

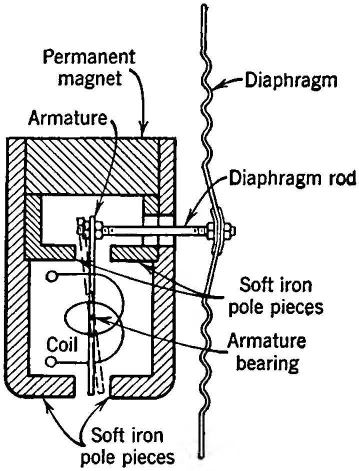


FIG. 6. Schematic diaphragm of a sound-powered telephone transmitter. The internal impedance is 900 ohms at 1000 cycles. For very loud talking the generated voltage is about 50 millivolts. This device may also be used as a receiver. (Courtesy Automatic Electric Co.)

A schematic diagram of a sound-powered magnetic telephone is shown in Fig. 6. Sound waves striking the diaphragm cause it to move, and this motion is imparted to the small soft-iron armature which conducts magnetic lines of force between the soft-iron pole pieces. As this armature moves in accordance with the speech sounds, the magnetic lines of force linking the coil are caused to vary, and this induces in the coil an electromotive force that is a good replica of the sound waves. This transmitter can be used also as a telephone receiver.

The Double-Button Carbon-Granule Microphone. This microphone has been extensively used in radio-broadcast and sound-amplifying systems. It is still used where high output is important. A cross section of a typical microphone is shown in Fig. 7.

The diaphragm of this microphone[11] is of Duralumin, 0.0017 inch thick, and is clamped securely at the outer edge to prevent slipping. The portions of the diaphragm in contact with the carbon granules are covered with a film of gold to ensure a low-resistance contact. The buttons are cylindrical and provided with a paper seal to prevent carbon leakage.

Connections for a microphone of the double-button type are shown in Fig. 8. The current passes through the two buttons in series, each of which has a resistance of about 200 ohms, giving an output impedance of 400 ohms resistance. When no sound waves excite the diaphragm, the resistance is constant, and, since the primary currents are equal

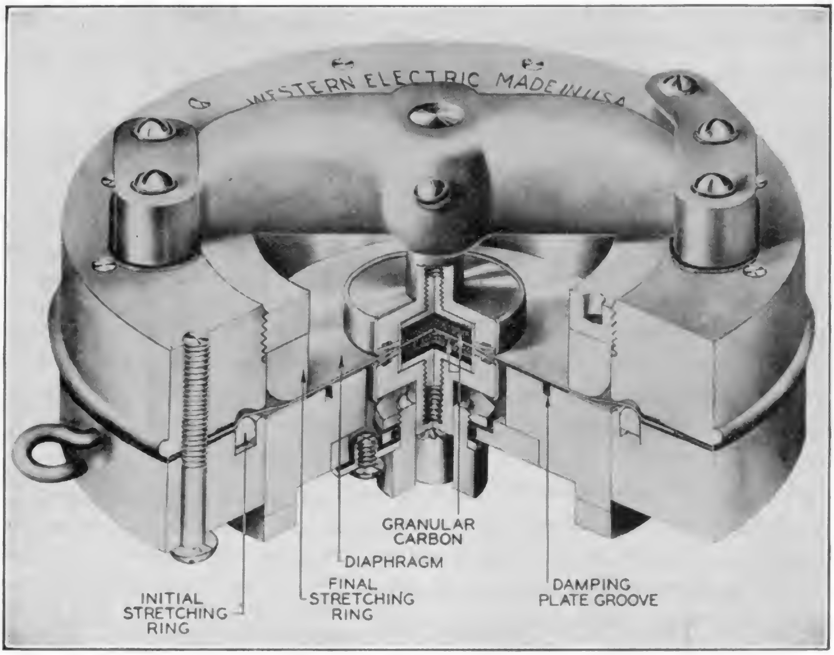


Fig. 7. Cross-sectional view of a Western Electric, double-button, carbon-granule microphone. (Courtesy Bell Telephone System.)

and in opposite directions, no magnetization is produced in the core of the transformer. When sound waves strike the diaphragm, however, the current is increased in one button and decreased in the other and a voltage is induced in the transformer secondary.

The diaphragm is stretched tightly and has a high natural frequency. Furthermore, as Fig. 7 indicates, the diaphragm is very close to a grooved metal plate which provides an air-damping chamber. This construction tends to make the output independent of frequency (Fig. 9), but this high quality is obtained at a sacrifice of sensitiveness. It has been stated[12] that the output of microphones of this type is *about* 10^{-8} watt and that this is of the same order as the speech power picked up by the microphone. In other words, the microphone does not amplify the speech power received by it from the sound waves like the

transmitter used in regular telephone service. The output is, nevertheless, much greater than that of any other of the microphones to be considered in the following pages. This is a decided advantage in many instances.

Typical current values are 25 milliamperes per button. The current should always be reduced to about zero value before opening the battery circuit. If this is not done, the high voltage induced in the transformer windings when the switch is opened may cause the carbon granules to arc across and thus injure the microphone. A filter system consisting of a 0.0014-henry choke coil in series with each line wire and a 0.02-microfarad condenser directly across each of the two transformer primary windings is sometimes used to reduce arcing.

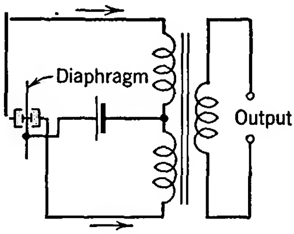


Fig. 8. Circuit for a double-button carbon-granule microphone.

The Condenser Microphone. This consists essentially of a fixed plate and a thin, tightly stretched diaphragm, the two being insulated and separated by about 0.001 inch. The basic circuit is shown in Fig. 10. When sound waves strike the diaphragm it moves and the capacitance varies causing an electric current corresponding to the sound waves to flow through the high resistance. The resulting voltage

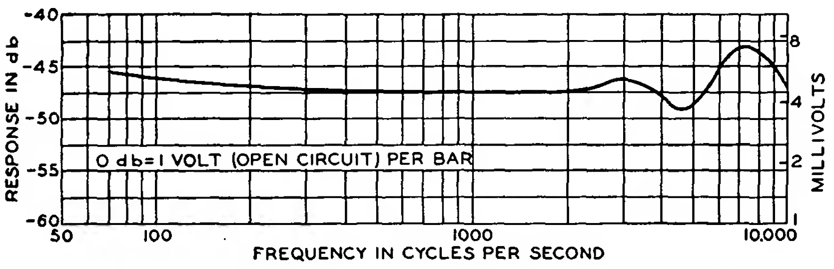


Fig. 9. Pressure calibration for the Western Electric microphone of Fig. 7. One bar is sound pressure of one dyne per square centimeter. For usual speech, the average pressure is about 0.4 bar, and for music it often is much higher, depending on the musical selection, etc.

drop is then amplified. The direct polarizing voltage is about 200 volts. The action of condenser transmitters and receivers was studied[11] by Dolbear in 1881. These microphones became of practical importance because of the improvements made by Wente and the development of vacuum-tube amplifiers.[13, 14] It has been stated[15] that modern acoustics began with the development, by Wente, of this microphone.

A condenser microphone is made small to minimize distortion of the sound waves; the diaphragm is about 1 inch in diameter. The capacitance is about 50 micromicrofarads. This requires (*a*) that very short leads must be used between the microphone and the amplifier and (*b*) that the input or grid resistor of the amplifier must be of high resistance. In a typical amplifier this is 100 megohms. Free-field response curves of the microphone and amplifier are shown in Fig. 11. Pressure calibration curves (page 110) are also available.[16] The output impedances of the amplifier commonly are 25 to 50 and 150 to 250 ohms.

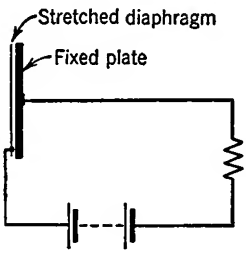


FIG. 10. The basic circuit of a condenser microphone. Although this microphone has a source of voltage in series, its output is very low.

The Moving-Coil or Electrodynamic (Dynamic) Microphone. In 1877 it was suggested[17] that the performance of Bell's magnetic transmitter could be improved by making the diaphragm of non-magnetic material and attaching to this diaphragm a light coil of wire which would vibrate in a magnetic field and generate a voltage in accordance with the sounds striking the diaphragm. Such a microphone was impracticable until amplifiers were available. Moving-coil microphones became commercially available about 1930.

Cross-sectional views of a moving-coil microphone are shown in

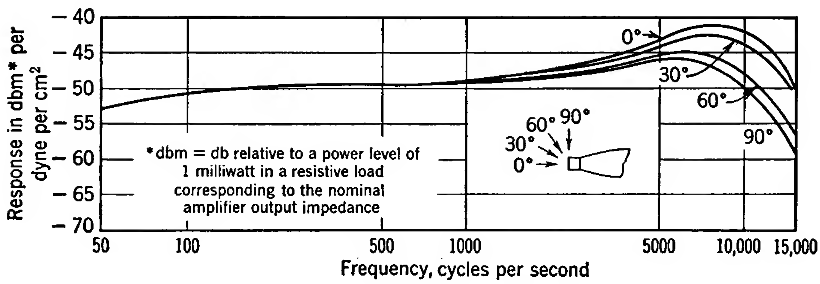


FIG. 11. Free-field response curves of a typical condenser microphone *and* associated amplifier. Arrows indicate direction of arrival of sound waves. (Courtesy Western Electric Co.)

Fig. 12. The diaphragm is made of Duralumin and is dome shaped to stiffen the center and ensure that the diaphragm moves with piston action. The moving coil fastened to the diaphragm consists of an aluminum ribbon wound on edge. The strong radial magnetic field

across the air gap in which the coil is located is produced by a permanent magnet.

As indicated by the arrows in Fig. 12 there are two major air chambers in this microphone. These chambers and the narrow connecting air slits improve the frequency response of the microphone.[18, 19]

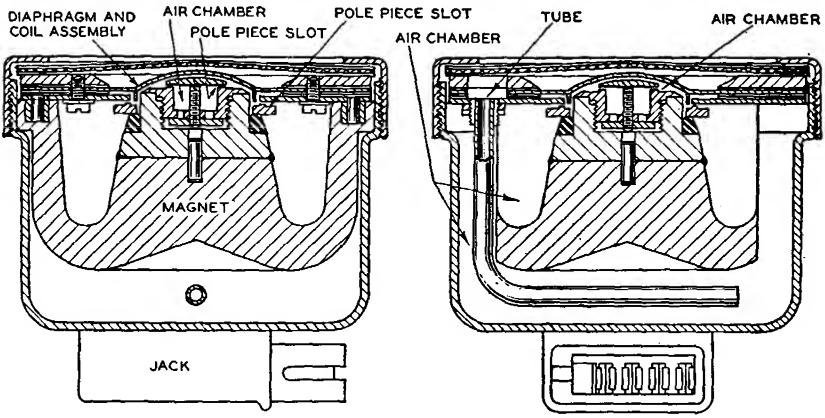


FIG. 12. Cross sections of a moving-coil or dynamic microphone. (Courtesy Western Electric Co.)

The shape of the housing is important in determining the directional characteristics. To obtain a non-directional microphone, a spherical housing was used in one type.[20] Frequency response curves of a typical dynamic microphone[21] are shown in Fig. 13.

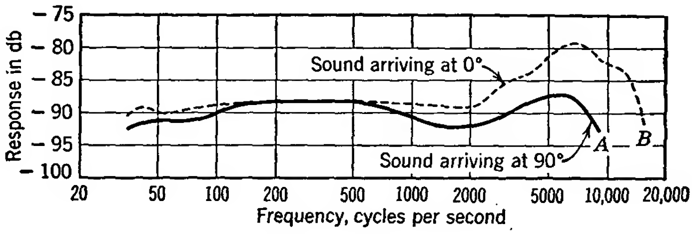


FIG. 13. Free-field response curve for a typical dynamic microphone. 0 decibels = 1 volt per dyne per square centimeter (open-circuit voltage across output impedance of 20 ohms). (Courtesy Western Electric Co.)

Several hundred feet of cable may be used between the moving-coil microphone and the associated amplifier because the impedance of the microphone is only about 20 to 40 ohms. The condenser micro-

phone element is, relatively speaking, a high-voltage low-current device of high internal impedance. The dynamic microphone element is (again speaking relatively) a low-voltage high-current device of low internal impedance. Shunting a condenser microphone with appreciable cable capacitance is objectionable but is permissible with dynamic microphones.

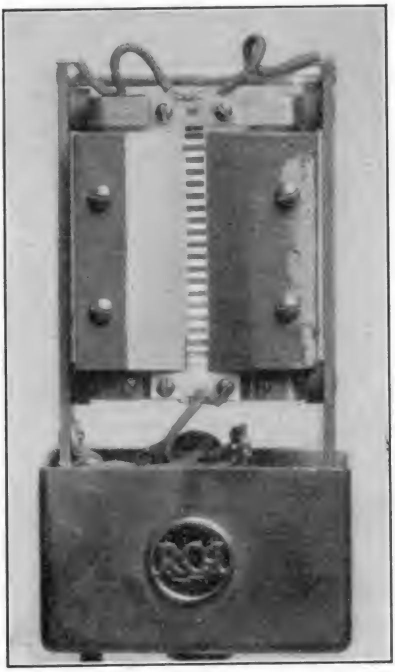


FIG. 14. A ribbon, or velocity, microphone. Output impedance from secondary of transformer, 50, 250, and 15,000 ohms. Output voltage 1425 $\times$ 10^{-6} volt open circuit across 250-ohm transformer taps with sound pressure of 10 dynes per square centimeter. (Courtesy Radio Corporation of America.)

The Ribbon or Velocity Microphone. The microphones considered previously are sometimes classified as **pressure-operated microphones** because they are enclosed in a housing and have a diaphragm, only one side of which is exposed to the sound waves. When sound waves strike the diaphragm, the difference in air pressure between the exposed and the enclosed sides causes the diaphragm to move. In the ribbon, or velocity, microphone there is no diaphragm; also the microphone housing is open to sound waves.

The ribbon of Fig. 14 is of thin, corrugated aluminum alloy and is loosely held between the poles of a strong permanent magnet. Sound waves approaching from the front, in a direction at right angles to the plane of the ribbon, can reach the back of the ribbon by flowing through the small slits between each side of the ribbon and the adjacent pole piece, or by flowing around the pole pieces to the back side.

The ribbon is moved by a difference in pressure, or *pressure gradient*. The difference in pressure between the two sides of the ribbon is caused by a difference in phase of the sound waves on the two sides of the ribbon.[15] These phase differences are caused by the difference in the length of the acoustic path from one side of the ribbon to the other side. Because the operation of the ribbon microphone depends on

air-particle motion and *velocity*, it is often called a **velocity-operated microphone.**

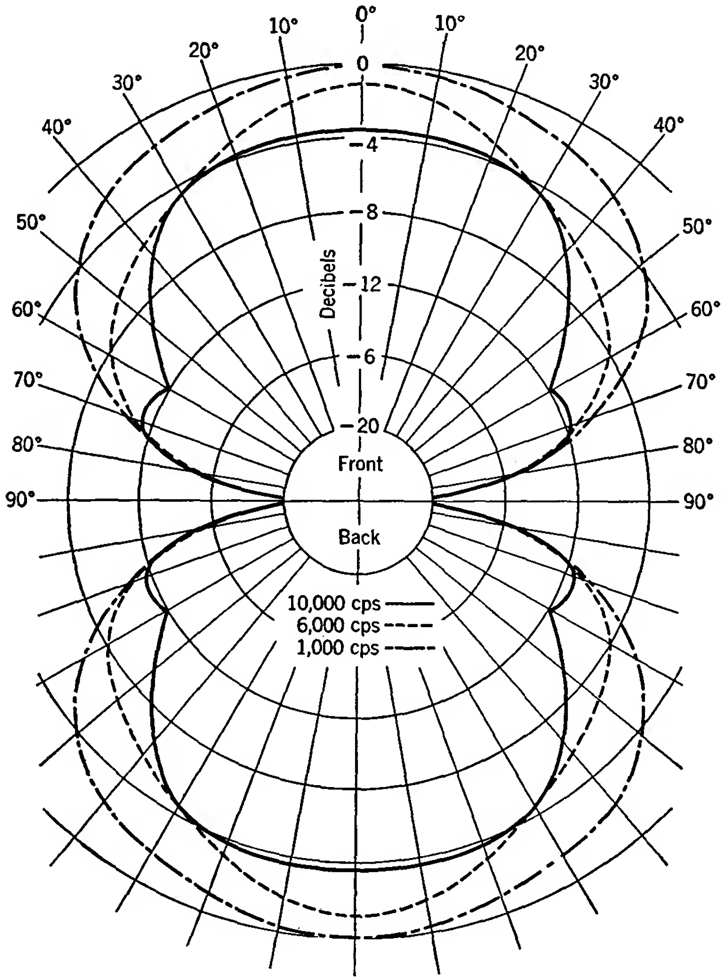


FIG. 15. Directional characteristics of a typical ribbon, or velocity, microphone. Front and back refer to the microphone which is assumed to be at the center of the figure. (Courtesy Radio Corporation of America.)

The motion of the ribbon in the magnetic field causes a signal voltage to be induced in the ribbon. Both the voltage and the internal impedance are low. These are increased by a transformer located beneath the pole pieces. Transformer output impedances are either 50 or 250 ohms.

The ribbon or velocity microphone is directional, as shown in Fig. 15. A sound coming parallel to the plane of the ribbon will produce no pressure difference on the two sides and accordingly no motion. This directional characteristic may be very useful for certain work such as sound pickup in highly reverberant rooms or in the presence of noise, especially if it is localized. By suitably enclosing one side of the ribbon microphone, it can be made directional.

Fig. 16. Parts of a diaphragm-type crystal microphone, and assembled microphone at lower right. The output is about −50 decibels, where 0 decibels is 1.0 volt per dyne per square centimeter. This microphone works into a 5-megohm load resistance. (Courtesy Brush Development Co.)

The Crystal Microphone. Certain substances, such as crystals of quartz or crystals of Rochelle salt, develop a potential difference between opposite surfaces when the crystals are deformed by mechanical forces. This is known as a **piezoelectric effect.**

Rochelle salt crystals are used in crystal microphones utilizing the piezoelectric effect. These crystals produce a larger potential difference than quartz. The Rochelle salt crystals are cut into thin slabs. Suitable metallic electrodes are provided on the faces of a slab. Wire leads are attached, and the element is sealed to prevent the entrance of moisture.

Two principles are used in making crystal microphones. One type makes use of the fact that, if sound waves strike the crystal element, these feeble waves are of sufficient strength to deform the crystal, thus producing a potential difference that corresponds to the impinging sound waves. This is the "sound cell" type. The "diaphragm type" of crystal microphone uses a diaphragm to intercept the sound waves, and the motion of the diaphragm is transmitted to the crystal which is

mounted so that motion of the diaphragm bends the crystal element (Fig. 16). Sound-cell crystal microphones have been built in which a number of crystals connected in series are employed, and sometimes they were connected in series-parallel combinations to give both increased output and low internal impedance.

With microphones and other devices using Rochelle salt crystals, the operating temperatures recommended by the manufacturer should not be exceeded or the crystal may be damaged permanently.

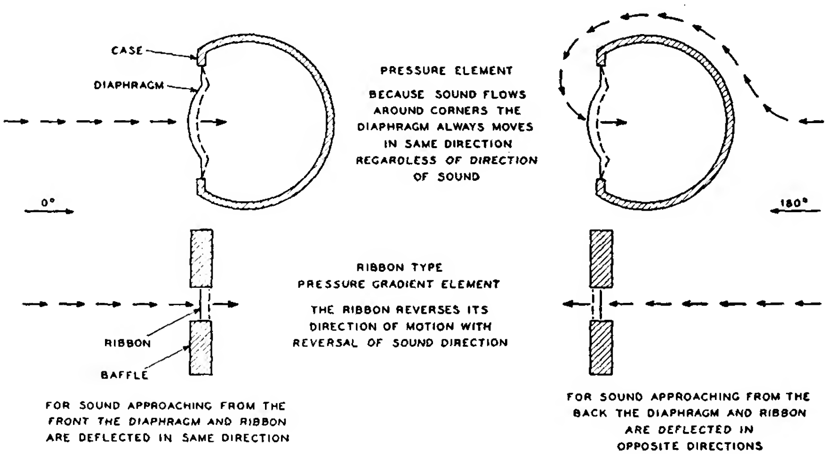


FIG. 17. Illustrating the principles of operation of the Cardioid microphone. (Courtesy Western Electric Co.)

Combined Pressure and Velocity Microphone. With the pressure-operated microphone, a sound wave coming from the front will force the diaphragm in. Similarly, a sound wave coming from the back of the microphone will flow around to the front with but little distortion and will also force the diaphragm in. These relations are shown in Fig. 17.

For the velocity microphone, however, sound waves from one direction cause a motion of the ribbon in one direction, and *corresponding* sound impulses from the *opposite* direction cause a motion in the *opposite* direction. The pressure or diaphragm-type microphone and the velocity or ribbon-type microphone can be used together to produce a highly directional sound pickup combination, as will now be explained.[22]

If the coil of wire attached to the diaphragm shown in Fig. 17 is properly connected *in series* with the ribbon of this same illustration, then, for sound approaching from the front, the voltage outputs of

the two generating elements will add, but, for sound coming from the back, the two generated voltages will subtract. The result is a directional sensitivity pattern that is heart shaped.

Microphone Calibrations. High-quality microphones, often of the condenser type, are used to measure the intensity of sound waves. Various methods of calibrating microphones have been summarized in references 15, 23, and 24.

There are two types of calibration, the **constant-pressure** or **pressure** calibration, and the **constant-field** or **field** calibration. The difference between the two is this: The microphone itself by reason of its presence in the sound field causes a distortion of the oncoming sound waves, although this effect is small for some types. Accordingly, a calibration made where the pressure is uniform over the diaphragm and measured at the diaphragm will not agree (especially at the higher frequencies) with a calibration made where the sound is picked up in an unobstructed space some distance from the source.

A thermophone sometimes is used in pressure calibrating microphones. It consists essentially of an enclosed chamber which can be tightly sealed against the face of the microphone to be calibrated. There are two very thin gold-leaf thermal elements near the bottom of the chamber. These are kept heated by a constant current, upon which an alternating current of the frequency at which the calibration is desired is superimposed. Gold leaf has low thermal capacity, and accordingly the impressed alternating current produces relatively large temperature variations. These in turn cause expansion and contraction of the surrounding gas, which constitute sound waves of determinable pressure. Calculations for determining this pressure can be made from the constants and operating data.[11]

Telephone Receivers and Loudspeakers. A **telephone receiver** is defined[1] as "a device whereby electric waves produce substantially equivalent sound waves." A **loudspeaker** is defined[1] as "a telephone receiver designed to effectively radiate acoustic power for reception at a distance."

Any electroacoustic transducer consists of an electric portion and an acoustic portion, although in some instances these two portions may be common. For instance, in some crystal microphones the sound waves strike the crystal (acoustic portion) and the crystal (electric portion) generates electric signals. Somewhat similarly, the received electric signals cause the iron diaphragm (electric portion) of the common telephone receiver to move back and forth, and this iron diaphragm (acoustic portion) thereby radiates sound waves.

The electric portion is the **motor element,** defined[1] as "that portion of a telephone receiver which receives power from the electric system and converts it into mechanical power." The acoustic portion is the **acoustic radiator,** defined[1] as "that portion of an electroacoustic transducer which initiates the radiation of sound vibrations." A telephone receiver or loudspeaker is a reciprocating electric motor loaded with an acoustic radiating system.

Types of Motor Elements. Many motor elements have been developed, and most of them can be classified under the following headings.

Condenser Motor Element.[25] As is well known, mechanical forces exist between the plates of a charged capacitor. Thus, if a special capacitor is constructed with one or more movable plates, if these plates are arranged so that they will radiate sound effectively, and, if speech or program electric voltage waves are impressed on the device, sound waves will be radiated.

Condenser receivers and condenser loudspeakers have not been used extensively. They are more delicate and more expensive, and they require higher voltages than other types. Also, condenser driving elements require a polarizing direct voltage to prevent the radiation of sounds of twice the frequency of the received electric signals.

Piezoelectric Motor Element. Electric signals impressed on the crystal electrodes cause the dimensions of the crystal to change in accordance. The crystal may radiate the sound waves directly or may be coupled mechanically to an acoustic radiator, such as a diaphragm or a paper cone.

Crystal telephone receivers are rugged, light in weight, sensitive, and have an excellent frequency response. The input impedance of one type is about 80,000 ohms at any frequency, and the sensitivity is 1.5 bars per volt at 1000 cycles.* Piezoelectric motor elements have been used to a limited extent in small loudspeakers.[26]

Magnetic Motor Element. This classification includes most of the motor elements used in modern telephone receivers and loudspeakers. Largely of historical interest are the **induction magnetic motor element**[27] and the **magnetostriction motor element.** Of practical importance is the **magnetic-armature motor element,** a device the operation of which involves vibration in some part of the ferromagnetic circuit.[1] Also of practical importance is the **moving-coil motor element** in which the mechanical forces are developed by the interaction of the field set up by the currents in the conductor and the polarizing field surrounding it.[1]

* Advertising literature of the Brush Development Company.

Magnetic-Armature Motor Element. From the general nature of the definition previously given, it follows that many magnetic-armature motor elements are possible. One type that was used early in radio is shown in Fig. 18. The armature is of soft iron and is held at the center so that it can move by flexing the support that holds it. The coil carrying the signal currents does not move.

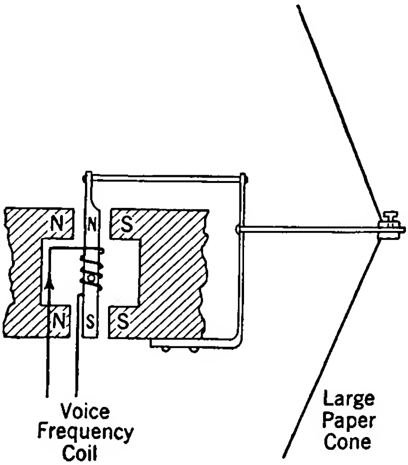


FIG. 18. Cross section of a balanced-armature cone loudspeaker.

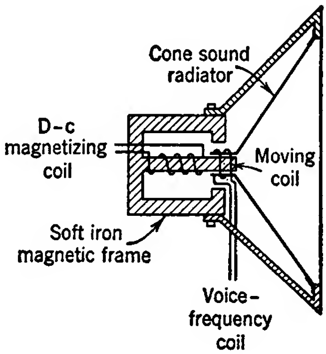


FIG. 19. Cross section of a moving-coil, or dynamic, loudspeaker. The magnetic field is often furnished by a permanent magnet.

Suppose that at a given instant the alternating signal current has the direction indicated. The soft-iron armature will then have the polarity shown, and the top of the armature will move to the right and the bottom to the left. On the next half cycle the current and the motion will be reversed. Because of this mode of operation, the device is often called a **balanced-armature motor element.**

In an early radio loudspeaker known as the **cone loudspeaker** the motion of the armature was transmitted to a large double cone, only a portion of the front cone being shown in Fig. 18.

Moving-Coil (Electrodynamic or Dynamic) Motor Element. This motor element is used almost universally in loudspeakers. A cross section of a dynamic loudspeaker using such a motor element is shown in Fig. 19. It has been pointed out[28] that a loudspeaker of this type was invented by Lodge in 1898.

The moving-coil motor element consists of a voice coil of a few

turns of wire suspended in a very strong, constant magnetic field. The coil is free to move back and forth axially. It is attached to a suitable acoustic radiator, such as a paper cone as in the radio dynamic loudspeaker, or a metal diaphragm, as in the driving unit used with large horn-type loudspeakers. The signal current variations in the voice coil react with the constant magnetic field and cause the coil and acoustic radiator to move and radiate sounds.

The strong constant magnetic field is produced by an electromagnet as shown in Fig. 19, or by a permanent magnet. If an electromagnet produces the field, the coil is sometimes made of a few turns of fairly heavy wire, and the direct current is furnished by a storage battery or by rectifiers. Or the coil can be made of a large number of turns of fine wire, and the exciting current is furnished by a rectifier supplying, perhaps, 50 milliamperes at several hundred volts. In many radio receiving sets the loudspeaker field coil also serves as the inductor, or "choke," in the filter of the power supply.

The impedance of the voice coil in a moving-coil driving unit is very low, a typical value being $Z = 8.8 \underline{/+25°}$ ohms at 1000 cycles, and a direct-current resistance of 5 ohms. Many moving-coil loudspeakers have an impedance-matching transformer (page 68) mounted on them to increase their impedance so that they will match amplifiers with high-impedance outputs. When the impedance of the voice coil was measured through an inexpensive transformer of this type, it was found to be $Z = 2400 \underline{/+30°}$ ohms. The difference in these phase angles is caused largely by the transformer.

As mentioned earlier in this section, the moving coil of a motor element used to drive a large horn is connected to a small metal diaphragm. A motor element of this type is shown in Fig. 20. The design of the diaphragm, air chamber, throat, and other details of a moving-coil motor element of this type is a specialized subject of much importance.

Types of Acoustic Radiators. The acoustic radiator is the portion of a telephone receiver or loudspeaker that initiates the radiation of sound vibrations.[1] Two types are commonly used: the small diaphragm and the cone. Fundamentally, they are the same, but practically they are quite different, particularly in size.

Diaphragms. Diaphragms are used in the receivers of common telephone sets and will be considered on page 119. Diaphragms are also used in the moving-coil motor elements for horn-type loudspeakers considered on page 114.

Cones. A **double cone,** with the driving element inside, was used with the type shown in Fig. 18. The cone used with the element of

Fig. 19 moves back and forth, even at the outer edge and is sometimes called a **free-edge cone.** These cones have been made of various substances, particularly paper and similar materials, and sometimes metal.[29] Surprisingly wide and uniform frequency response is obtained with a loudspeaker of this type, as indicated in Fig. 21.

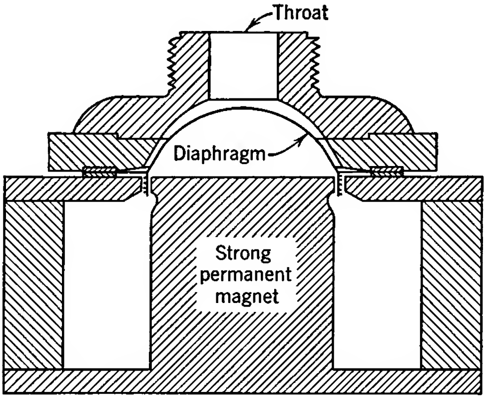


FIG. 20. Cross section of a driving unit for a horn-type loudspeaker. The impedance is about 15 ohms at 1000 cycles, and is largely resistance, a 15° lagging angle being typical. This unit, when coupled to a horn may be driven with approximately 20 watts. The efficiency is about 35 per cent. (Courtesy Racon Electric Co.)

Dynamic Loudspeakers. This type is used almost exclusively in radio. It consists essentially of a moving-coil motor element driving a free-edge cone. The cones used are of various types and shapes and are made so that they move approximately with piston action.

A dynamic loudspeaker should be mounted in a cabinet or in a **baffle** if it is to operate satisfactorily, particularly at low frequencies. When the cone moves out, a condensation is produced on the front side and a rarefaction occurs on the rear side. The air accordingly flows around the edge, neutralizing the pressure difference

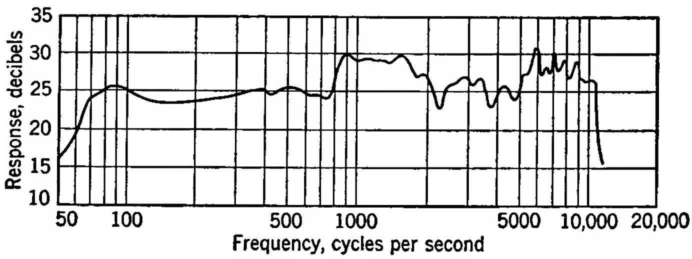


FIG. 21. Response curve of a moving-coil loudspeaker with a thin metal cone radiator. (Reference 29.)

and hence largely preventing sound radiation at that frequency. The baffle corrects this by providing a long path from the front of the diaphragm around to the back, thus preventing neutralization until

after considerable sound has been radiated. The dimensions of the baffle should be such that the shortest air path between the front and the back of the cone is at least one-fourth the wavelength of the lowest note to be reproduced. Thus, for a 50-cycle note the wavelength would be approximately $(1125 \times 12)/50 = 270$ inches. One-fourth of this is 67.5 inches or about 5.6 feet for the distance from the front around the baffle to the back of the cone.

From the standpoint of frequency response alone, a dynamic loudspeaker mounted in the center of a large wall would be ideal. However, since much sound is radiated from the back of the cone, this would result in low efficiency. Dynamic speakers are often mounted in a cabinet, the back of which is left open. The cabinet then acts as a baffle and greatly influences the radiated sound. When it is necessary to enclose the back of the cone entirely, the box should be as large as possible and should be lined with hair felt or other sound-absorbing material to prevent the box from causing serious distortion. Distortion is caused in two ways: *First*, the reflected waves from the unlined walls strike the cone and interfere with the proper operation of the speaker, thus impairing its output. *Second*, the box is mechanically resonant for certain frequencies and if unlined is more likely to be set in undesired resonant vibration.

Horn-Type Loudspeakers. Horn-type loudspeakers are used extensively where large audiences are to be served, as in a large auditorium or in a stadium. These loudspeakers usually consist of a horn attached to a moving-coil driving unit such as is shown in Fig. 20. The function of the horn is interesting as the following quotation[30] indicates.

Contrary to the prevalent conception, the horn does not merely gather up the sound energy from the receiver and concentrate it in certain directions. Its relation to the diaphragm is much more intimate. It causes an actual increase in the load on the diaphragm, making it advance against a greater air pressure, and withdraw from a greater opposing rarefaction. Anyone can assure himself that the average sound energy in a room is greatly reduced on removing the horn from a good loud speaker. And frequently when the horn is removed the amplitude of vibration of the diaphragm becomes so great that it strikes against the pole pieces. A receiver element without a horn is analogous to a motor without a connected load; or better yet, a receiver element without a horn is like a closed oscillation circuit from which little radiation takes place (radiation resistance zero); while with a horn it is like an open oscillation circuit with an antenna (radiation resistance considerable). The horn is the antenna of the loud speaker.*

* Reprinted by permission, courtesy C. R. Hanna, J. Slepian, and the American Institute of Electrical Engineers.

The Air Chamber. The horn itself is connected acoustically to the diaphragm by the throat and air chamber as indicated by Fig. 20. This air chamber acts as an acoustic transformer for, owing to the differences between the area of the diaphragm and the area of the throat, a small diaphragm velocity gives the air in the horn a greater velocity and much higher air pressure.[30]

To explain further the theory of the air chamber and the throat, assume that the volume of the chamber is so small that as the diaphragm is moved forward all the air is forced out instead of being compressed. If the area of the throat is now reduced, it is apparent that, as this area becomes smaller, the mechanical load on the diaphragm grows larger. Thus, if the throat is closed, the diaphragm will be damped (page 123), and substantially no motion will be possible when the coil is energized. The size of the area of the throat is made such that the diaphragm is effectively coupled to the air, and the opposition to motion is then almost a pure mechanical resistance.[1] For good frequency response, this resistance relation should hold over the entire frequency range. The area of the throat should not be made too small or air friction will be excessive.

Size of the Horn Mouth. If the area of the horn mouth is not correct, sound waves of the lower frequencies will not be effectively radiated from the horn but will be reflected back down the horn.[30] As the sound waves suddenly leave the mouth of a horn, the waves greatly increase in volume and decrease in pressure. If the air pressure outside the mouth is lower than that immediately within, the air velocity just within the mouth will be increased. This in turn will cause the pressure just behind to decrease and the velocity to increase. By this action, therefore, sound waves are propagated back down the tube to the diaphragm.

These reflected sound waves represent acoustic power which is not radiated into the air. Also, these reflected waves either aid or oppose the action of the diaphragm, depending on the phase relations they have upon reaching the diaphragm, and hence on their wavelength. This action will result in distortion. Hanna stated[31] that such reflections will not be objectionable if the diameter of the horn mouth is greater than one-fourth of the wavelength of the lowest frequency that it is desired to radiate. In horn design, the mouth is given an area approximately equal to the area of a circle having a diameter equal to one-quarter the length of the wave of lowest frequency to be radiated; this applies approximately for rectangular openings.

Rate of Taper. From the preceding discussions it is seen that the area of the throat must be small to load the diaphragm properly, and

that the mouth of the horn must be large to radiate the lower frequencies into the air effectively. The next consideration is the shape and length of the horn to connect these two extremities properly.

Wave reflection will occur at any discontinuity. If reflection *along* the horn is to be minimized, the relative increase in cross-section area should be uniform. Thus, the **exponential horn,** "whose sectional area varies exponentially with its length,"[1] is used. This horn is defined[1] by the following relations:

$$S = S_o \epsilon^{Tx}, \tag{12}$$

where S is the area of plane section of the horn normal to the axis at a distance x from the throat of the horn; S_o is the area of plane section of the horn normal to the axis at the throat; and T is a constant which determines the rate of taper of the horn.

The effectiveness with which an exponential horn transmits sound energy to the mouth is determined by the frequency of the sound wave in relation to the rate of taper, or rate at which the horn opens out, that is, to T. It can be shown[30] that, for frequencies below about $f = 4000T$, the transmission is poor, and, if the frequency is further reduced, a cutoff frequency is soon reached. At high audio frequencies the propagation along an exponential horn is excellent. In the past, **conical horns** having poor characteristics were used.

Hanna gave a simple rule for laying out exponential horns.[31] In such horns the area doubles at equal intervals along the length. Since the cutoff frequency is a function of the rate of expansion, the cutoff is also a function of the length between two circles having a ratio of 2 to 1. As he pointed out, if the area doubles every 3 inches, the cutoff frequency will be 256 cycles; if it doubles in 6 inches, it will be approximately 128 cycles; and if every 12 inches, it will be 64 cycles.

Since the throat area must be small and the mouth opening large, and furthermore since good frequency characteristics demand that the *rate of taper* be not too large, the exponential horn must be comparatively long. In the open, or where space is not limited, a trumpet horn is sometimes used, the maximum length of such horns being about 6 feet. Where space requirements are important, horns are sometimes coiled, lengths as great as 12 feet having been obtained in this way. At present a folded-horn construction is usually employed. The material used in constructing a horn should be such that portions of the horn do not vibrate or rattle.

Short Exponential Horns.[15] The diaphragm of the driving unit of Fig. 20 sends sound waves toward the open mouth of the horn. The air pressure is high at the narrow throat and decreases in intensity as

the wave travels toward the mouth. If a short section of an exponential horn is coupled to a loudspeaker with a large diaphragm, such as Fig. 19 and if this large diaphragm produces the same acoustic pressure at a given point in the short horn as the driving unit of Fig. 20 produces at a corresponding point in a long horn, the performance should be similar.

Short exponential horns with loudspeakers of the general type of Fig. 19 having diaphragms about 6 inches or more in diameter are used

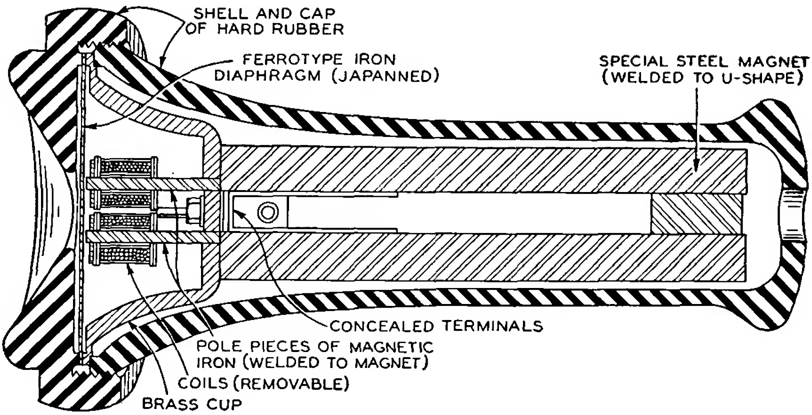


FIG. 22. Cross section of a Western Electric telephone receiver. This is used with the transmitter on page 95. This receiver, like the transmitter there described, will gradually be replaced by those of later design, but many of this older type will remain in service for years.

as just described. From the discussion given for the long exponential horn with a small throat, it is apparent that the efficiency would not be so high. The frequency response, however, may be excellent, and the directional characteristics are good.

In some instances the horn used with a loudspeaker such as Fig. 19 is very short, often merely a flared system of boards of a variety of shapes. Such arrangements are commonly called **directional baffles.** They increase the efficiency and directivity but little over that obtained with a flat baffle.

Hand and Head Receivers. A **hand receiver** is[1] a "telephone receiver designed to be held to the ear by the hand," and a **head receiver** is[1] a "telephone receiver designed to be held to the ear by a headband." Because their fundamental principle of operation is the same, no further distinction will be made between them. They will be referred to as **telephone receivers** or **receivers** in this chapter.

For illustrating the theory of operation, the old-style telephone receiver will be used because of its simplicity.

The permanent magnet of the receiver of Fig. 22 provides a constant magnetic field which passes from the north pole, through the magnetic (soft-iron) pole piece, across the air gap, through the soft-iron diaphragm, across the air gap, through the pole piece, and to the south magnetic pole. The coils through which the speech currents flow are placed on the soft-iron pole pieces and are connected in series so that they aid.

The necessity for the constant pull on the diaphragm is made clear by Fig. 23. Suppose that one cycle of alternating current passes through the coil in (*a*). As the current increases from zero to a positive maximum, the diaphragm is pulled in to the dotted position. The adjacent air particles on the right of the diaphragm *D* will flow in, thus causing a **rarefaction.** Now as the current dies out to zero, the diaphragm will return to a position of the zero displacement, and in so doing the air particles on the right will be compressed and a **condensation** will be produced. When the current builds up in the negative direction the diaphragm will be again drawn in, producing another *rarefaction;* and, when it again dies out to zero, the diaphragm will return to the position of rest and will produce another *condensation.* When *one* cycle of alternating current flows through the speech coils, *two* complete cycles of sound waves are set up. Thus, if a constant pull is *not* exerted on a diaphragm, the reproduced sound waves will be *twice* the frequency of the speech currents.

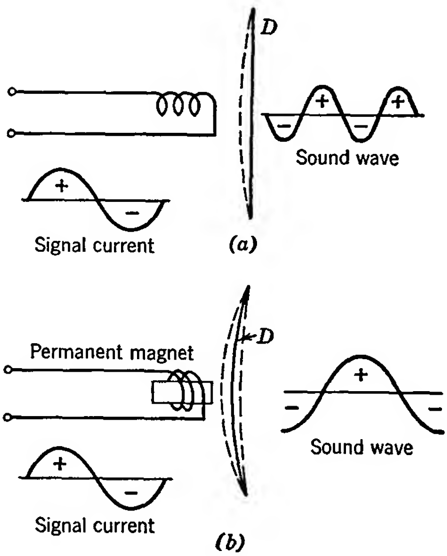


FIG. 23. Illustrating how a double-frequency tone is produced by a receiver not having a constant pull on the diaphragm.

If the windings are placed on a permanent magnet as in (*b*), the diaphragm is bowed in when no current flows, as shown by the full line *D*. When the current increases from zero to a positive maximum, the diaphragm is pulled in further to the dotted position, producing *one half* of a *rarefaction.* When the current dies out to zero the dia-

phragm returns momentarily to the position shown by the full line producing *one half* of a *condensation.* As the current builds up in the opposite direction the flux due to this current neutralizes part of the flux from the permanent magnet and the diaphragm moves to the outward dotted position, thus causing the *other half* of the *condensation.* As the current dies out to zero, the diaphragm returns momentarily to the full-line position, causing the *other half* of the *rarefaction.* With a constant pull on the diaphragm, therefore, *one cycle of current causes one cycle of sound wave,* and thus the frequency of the sound is the same as that of the exciting current.

Theory of Telephone Receiver Operation.[32, 33] The total magnetic flux ϕ crossing the air gaps between the pole pieces and the iron diaphragm is composed of the constant flux ϕ_o produced by the magnets, and the variable flux ϕ_i caused by the voice currents passing through the coils on the soft-iron pole pieces. That is,

$$\phi = \phi_o + \phi_i. \tag{13}$$

The air gaps tend to keep the total reluctance of the magnetic path independent of the current intensity, and it can therefore be assumed that the flux ϕ_i produced by a sine-wave test current is proportional to the current intensity, or

$$\phi_i = KI_m \sin \omega t. \tag{14}$$

Combining equations 13 and 14,

$$\phi = \phi_o + KI_m \sin \omega t. \tag{15}$$

As shown in most textbooks presenting magnetic theory, the force of attraction between two portions of a magnetic circuit separated by an air gap varies as the flux squared, and it can therefore be written that the force F acting on the diaphragm is

$$F = K_1\phi^2 = K_1(\phi_o + KI_m \sin \omega t)^2. \tag{16}$$

When this expression is squared it becomes

$$F = K_1\phi_o^2 + 2K_1K\phi_oI_m \sin \omega t + K_1K^2I_m^2 \sin^2 \omega t. \tag{17}$$

Since from trigonometry $\sin^2 \omega t = (1 - \cos 2\omega t)/2$, equation 17 can be written

$$F = K_1\phi_o^2 + 2K_1K\phi_oI_m \sin \omega t + \frac{K_1K^2I_m^2}{2} - \frac{K_1K^2I_m^2 \cos 2\omega t}{2}. \tag{18}$$

Equation 18 is of importance as it indicates the forces acting on the receiver diaphragm. These are: (1) a steady pull $K_1\phi_o^2$ produced by

the permanent magnets; (2) a force $2K_1K\phi_oI_m \sin \omega t$ proportional to the product of the *strength* of the permanent magnets and the value of the *instantaneous current* flow; (3) a force $(K_1K^2I_m{}^2)/2$ which is constant; and (4) a force $(K_1K^2I_m{}^2 \cos 2\omega t)/2$ which has a frequency $(2\omega t)$ *twice* that of the impressed current. Thus, part 4 produces a double-frequency sound, causing distortion. An examination of parts 2 and 4 will show that, for good quality, the flux ϕ_o from the permanent magnet should be made large so that the magnetic effect of the speech currents is *small* in comparison with that of the permanent magnet. Then, part 2 will be large and the volume will be sufficient, but part 4 will be small and there will be little distortion.

There is also a force acting due to the eddy currents induced in the diaphragm. The voice-frequency component of the flux passes through the diaphragm, and as this flux changes it induces eddy currents. It will now be shown that these eddy currents cause distortion.

Assume that a sine-wave current $i = I_m \sin \omega t$ flows through the coils of the receiver. There is little hysteresis owing to the large air gaps, and therefore the flux produced is in phase with the current. This varying flux will induce a voltage in the diaphragm proportional to the rate of change of flux, but lagging 90° behind it. The eddy currents flowing in the diaphragm will be in phase with the voltage and will accordingly lag 90° behind the currents producing them. Thus, with respect to the useful current $i = I_m \sin \omega t$, there will be eddy currents having the value $i_e = I_{e(m)} \sin (\omega t - 90)$.

The distorting force acting on the diaphragm because of the reaction of the eddy currents and the useful voice currents is proportional to the product of the two equations just written. That is,

$$F_e = K[(I_m \sin \omega t)(I_{e(m)} \sin (\omega t - 90))]. \tag{19}$$

This can be written in the form

$$F_e = KI_mI_{e(m)} (\sin \omega t) \sin (\omega t - 90). \tag{20}$$

Since from trigonometry $(\sin \omega t)$ $(\sin (\omega t - 90))$ is equal to $\frac{1}{2} \sin 2\omega t$, equation 20 becomes

$$F_e = \tfrac{1}{2}KI_mI_{e(m)} \sin 2\omega t. \tag{21}$$

Therefore, since $2\omega t$ is twice the frequency of ωt, the eddy currents in the diaphragm cause a double-frequency component and thus distort the original speech sounds.

This double-frequency effect can be illustrated by Fig. 24. The curve i represents the voice currents in the receiver windings, and i_e represents the induced eddy currents and the resulting flux ϕ_e in the receiver diaphragm, lagging by approximately 90°. Since the force on

the diaphragm is at any instant the product of the current and the flux, this force will be a double-frequency wave as indicated by the heavy line. Diaphragms of high-resistance material tend to prevent large eddy currents and thus minimize this double-frequency distortion.

The degree of saturation of the receiver diaphragm has an influence on distortion. Because of hysteresis, if the diaphragm is not operated fairly high on its magnetization curve, the increase and decrease of flux with current will not be proportional, and thus the diaphragm will not follow the current variations. Since the diaphragm is thin, the desired magnetic operating point is easily reached.

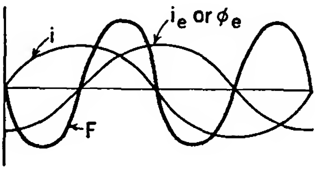


FIG. 24. The eddy currents in the diaphragm reacting with the voice current in the windings produce a double-frequency force F.

The air gaps between the pole pieces and the diaphragm tend to reduce distortion by making the overall magnetization characteristics approach a straight line. Then, the rise and fall of flux in the magnetic path will closely follow current variations instead of following a hysteresis curve.

Direct current passing through the windings of a receiver not designed for it may cause distortion by opposing the flux from the permanent magnet, and thus shifting the point of operation to a non-linear portion of the magnetization curve. Also, direct current in opposing the constant magnetic flux will weaken the field, making the receiver less sensitive. Furthermore, if the current is strong enough it may clamp the diaphragm against the pole pieces or even burn out the windings.

Distortion is also caused by the mechanical resonance of the receiver diaphragm. This causes a greater response at the resonant frequencies and causes the diaphragm to tend to continue to vibrate at these frequencies. The diaphragms of the new telephone receivers are designed to minimize these effects.

Acoustic distortion of the radiated sound waves is produced by the effects of the air cavities of the receiver and of the listener's ear.

Input Impedance of Telephone Receivers. A receiver or a loudspeaker is an electric motor, and information regarding operation can be obtained from input impedance measurements. The data and the discussion in the following pages are for a small low-impedance head receiver. This receiver was resonant at about 840 cycles. Similar tests can be made on other telephone receivers and can also be

made on loudspeakers. The results would be somewhat different because some modern receivers and loudspeakers are less resonant.

If a receiver is connected to an impedance bridge as in Fig. 25, if the test current indicated by the milliammeter is held constant by rheostat r, and if the resistance R and the inductance L required to balance the receiver impedance at different frequencies are measured, it will be found that these measured values vary greatly with test conditions. If the receiver is radiating sound energy into a small closed box the impedance will be greatly different from that if it is radiating into a large room. Or if the receiver is held to the ear the impedance will be different from that measured if it is radiating into a room. The impedance of a receiver or loudspeaker varies with the acoustic load on the diaphragm.

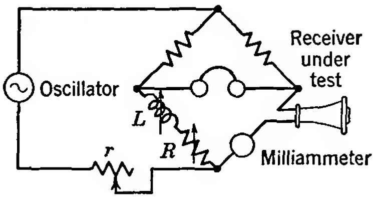


FIG. 25. Impedance bridge for measuring receiver constants.

The vector difference between the normal and the blocked impedance of an electroacoustic transducer such as a telephone receiver or a loudspeaker is defined as the **motional impedance**[1] of the device. This motional impedance is proportional to the back electromotive force induced in the receiver windings by the motion of the receiver diaphragm. The reluctance of the magnetic circuit of a receiver changes with the motion of the receiver diaphragm, and thus the flux linking the receiver coils varies, inducing a back electromotive force in these coils. If the diaphragm is damped, the diaphragm motion is influenced, and thus the induced back electromotive force and therefore the impedance are changed.

Receiver Characteristics with Diaphragm Blocked. The impedance is measured with the diaphragm held stationary in such a way *that the normal magnetic field relations will not be changed.* That is, the diaphragm is held in the same position as when the receiver coils are not carrying current. Several methods[33] have been developed for blocking the diaphragm. One method employs air damping and is accomplished by closing the hole in the receiver cap with an air-tight plug of wood or wax. This last method was used in obtaining the data here included.

Resistance and reactance curves obtained with the diaphragm of a telephone receiver blocked are as shown in Fig. 26. The test current for this figure and those that follow was 1.0 milliampere.

An impedance diagram corresponding to these readings is included in

Fig. 27. The line R_1 represents the direct-current resistance of the receiver windings. The line R_2 represents that part of the total effective resistance due to power consumed largely in hysteresis and eddy-current losses in the magnetic circuit. The line $R_1 + R_2$ is the

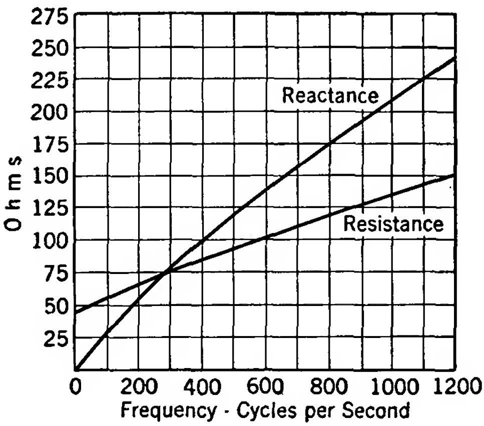


FIG. 26. Resistance and reactance curves with diaphragm blocked.

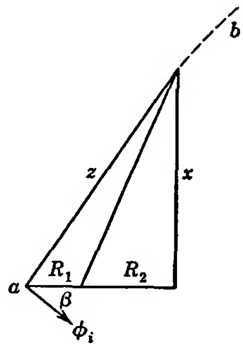


FIG. 27. Impedance diagram with diaphragm blocked.

resistance component, the line x the reactive component, and z the impedance of the receiver with the diaphragm blocked. The corresponding vector diagram is given in Fig. 28. The vector E_z represents the voltage across the receiver, leading the receiver current I by an angle α. The vector ϕ_i is the flux produced by the current I and lags I by the angle β because of the hysteresis and the skin effect in the magnetic circuit.[33]

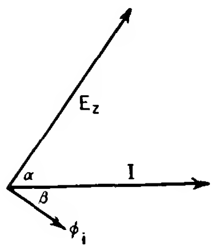


FIG. 28. Vector diagram for a receiver with diaphragm blocked.

The reactance of Fig. 27 increases almost directly as the frequency. The alternating-current portion R_2 of the damped receiver resistance is largely composed of hysteresis and eddy-current losses. Because of these losses the measured effective resistance varies as the frequency to some power greater than unity (page 52), and the point b of Fig. 27 will follow the approximate path a-b which bends continually to the right. This bending tendency is evident in Fig. 29, which represents the intersections of x and z (or the point b of Fig. 27) for a number of frequencies.

Receiver Characteristics with Diaphragm Free. The impedance characteristics with the diaphragm blocked are similar to those of a coil on an iron core containing an air gap. With the diaphragm free, however, the motional impedance and the mechanical resonance of the diaphragm greatly change the impedance characteristics.

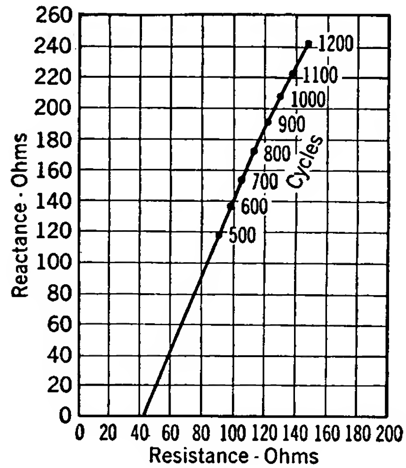


FIG. 29. Impedance curve with diaphragm blocked. Points indicate test frequency.

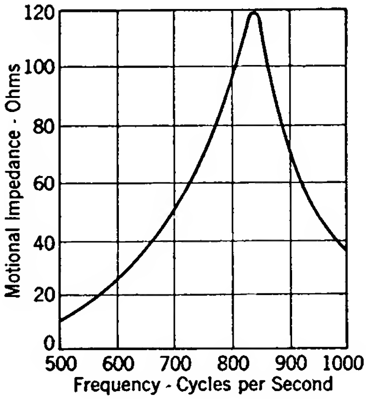


FIG. 30. Motional impedance curve for a telephone receiver; diaphragm free.

The receiver diaphragm has a certain mass which through its inertia effect tends to prevent changes in diaphragm motion. This causes an effect analogous to inductive reactance. The diaphragm is of elastic material and thus possesses compliance (displacement per unit force or reciprocal of stiffness[1]) which causes an effect analogous to capacitive reactance. At certain frequencies these two mechanical reactances neutralize each other, and mechanical resonance of the diaphragm occurs.

When mechanical resonance is reached, the motion of the diaphragm is greatest and is in phase with the impelling force. Since the greatest motion means the greatest flux change, the largest value of back voltage will be induced in the receiver windings and thus the motional impedance will be greatest at this point (Fig. 30).

The manner in which the total impedance changes with frequency is evident from Fig. 31. These curves show that at the resonant point the reactive component becomes very small and the resistive component becomes large, indicating that the diaphragm motion is in phase with the impressed electrical impulses.

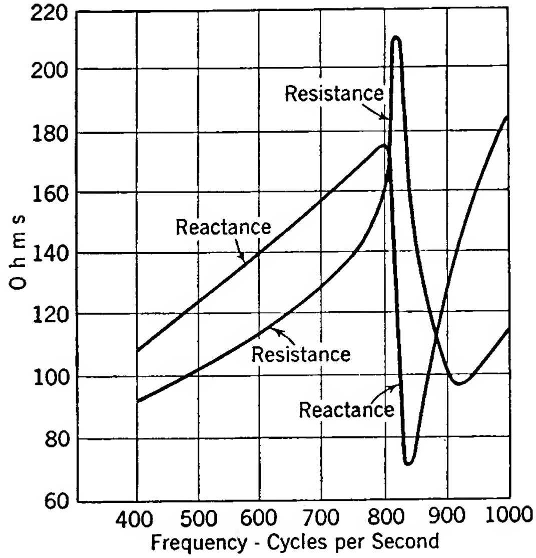


FIG. 31. Resistance and reactance curves for a telephone receiver with diaphragm free.

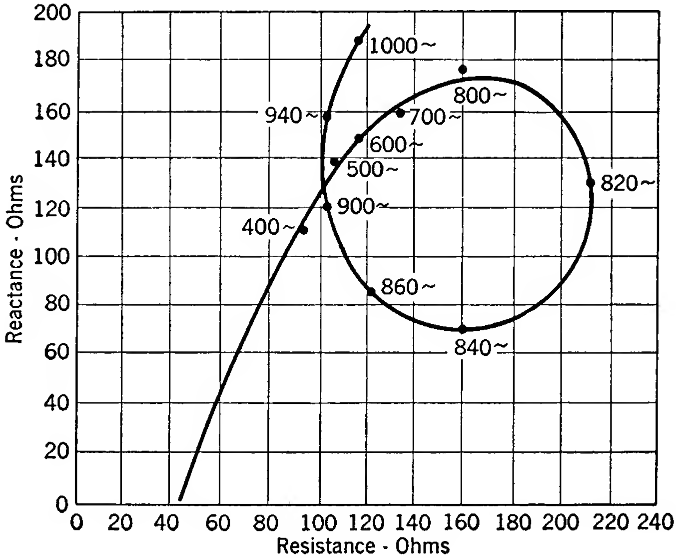


FIG. 32. Impedance curve for a telephone receiver with diaphragm free. Points indicate test frequency.

When the resistance and reactance components measured with the diaphragm free are plotted, the interesting curve of Fig. 32 (corresponding to Fig. 29) is obtained. Since Figs 29 and 32 are for the same receiver, they may be plotted together as in Fig. 33. The vector difference between corresponding points on these two curves as shown in Fig. 33 is the motional impedance.

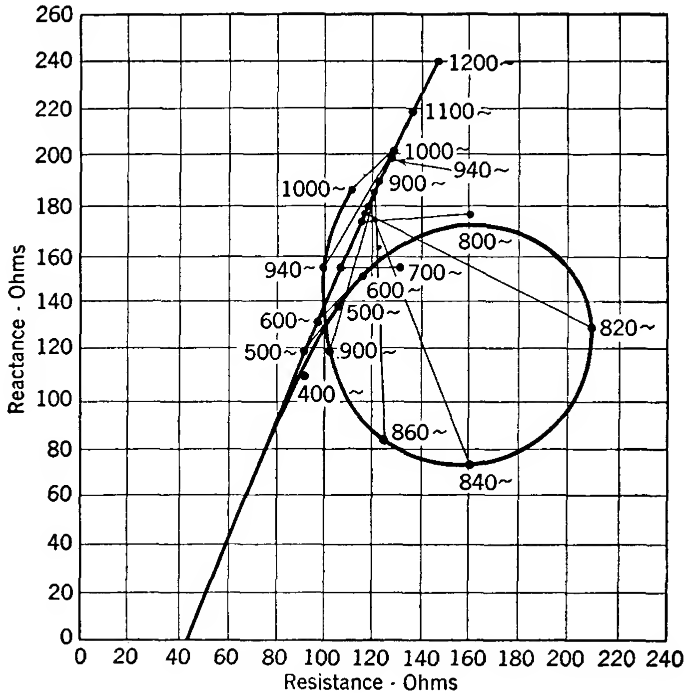


FIG. 33. Blocked and free impedance curves (Figs. 29 and 32).

If each of these motional impedance lines is plotted from a common point the curve of Fig. 34 will be produced. As is shown, the length of the motional impedance line at the resonant frequency f_r determines the diameter of the circle.

The vector diagram of a telephone receiver with the diaphragm free can now be drawn. As Fig. 35 indicates, this diagram is produced by adding the motional impedance circle of Fig. 34 to the impedance diagram of Fig. 27. This figure represents conditions at the resonant frequency, the line Z_f being the total measured impedance at this frequency. As the test frequency is varied the total impedance Z_f intersects the motional impedance circle at different points, and also the

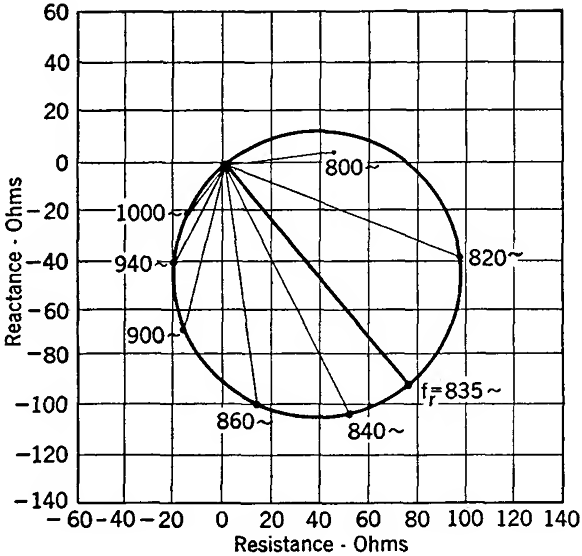


FIG. 34. Motional impedance diagram of a telephone receiver.

point b moves along the curve of Fig. 29. Thus, the curve of Fig. 32 is produced.

As indicated in Fig. 35 the motional impedance Z_m at the resonant frequency lags behind a line parallel to the X axis by the angle 2β, that is, by an angle twice as great as that by which the receiver flux in Fig. 27 lags the current through the receiver windings. This can be explained by an analysis of receiver action.[33]

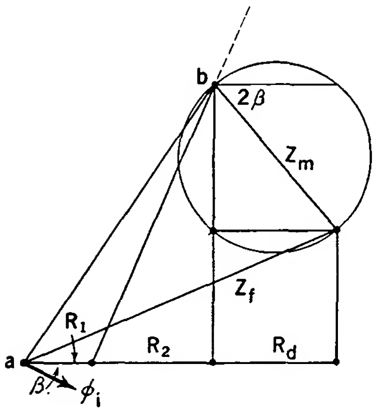


FIG. 35. Impedance diagram for a telephone receiver at the resonant frequency.

Although the discussion given in the preceding pages was for the old-type telephone receivers, which are highly resonant, attention again is called to the fact that the general theory applies also to the new-type receivers and to certain loudspeakers.

Impedance tests made on moving-coil loudspeakers with the diaphragm blocked, and free, and with and without the horn attached, and with different types of horns, provide interesting and useful data. Such tests can be performed by driving

the motor element with a good amplifier, and by measuring the current and voltage with thermocouples and vacuum-tube voltmeters, and the phase angle with a cathode-ray oscilloscope.

Receivers for Telephone Handsets. The receivers considered in the preceding pages are becoming obsolete. A serious objection is the resonant diaphragm. This resonance was desirable in early receivers because it increased the sensitivity. This is no longer necessary because telephone circuits are better designed and vacuum-tube amplifiers are available for long-distance service.

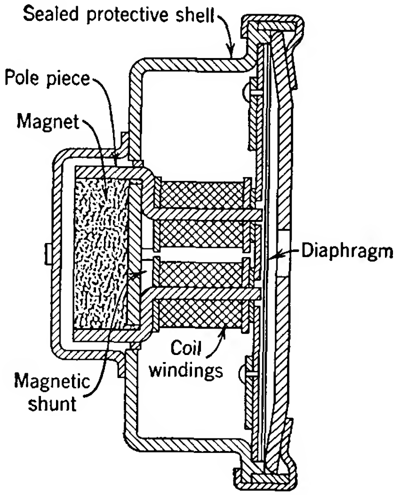


FIG. 36. Cross section of a typical telephone handset receiver. (Courtesy Automatic Electric Co.)

The receivers used for modern handsets are of the capsule type and are made so that they cannot be dismantled in the field. The basic principle of operation is the same as for the receivers considered previously. Methods of construction used and the improvement in frequency response are shown in Figs. 36 and 37.

A modern receiver has a small, yet powerful, permanent magnet which is often made of Alnico. The diaphragm is made of good flux-conducting material. Also, the diaphragm is constructed and acoustically damped so that resonances are minimized.[34, 35]

The resonant diaphragms of the old receivers produced loud clicks when excited by transients such as those caused by switching. These clicks sometimes caused **acoustic shocks** that were bothersome and occasionally injurious. Such disturbances are not so pronounced with non-resonant receivers.

Miscellaneous Receivers and Loudspeakers. In addition to the receivers and loudspeakers considered in the preceding sections, many other types have been developed.[5, 32, 36]

Thermal Receiver. The **thermal receiver** or **thermophone** is seldom used as a receiver, but it is employed to calibrate transmitters. A thermal receiver to be inserted into the ear for receiving telephone or radio messages was devised in 1906 by Eccles;[5] it is shown in Fig 38.

Pneumatic Loudspeakers.[1] A device called a Stentorphone was invented by Gaydon.[5] An air valve controlled by the electric signals to be reproduced regulated the flow of compressed air into a horn. A modern version is the Vocal-Aire loudspeaker,[37] a highly successful device for producing very intense sounds. This has a good response

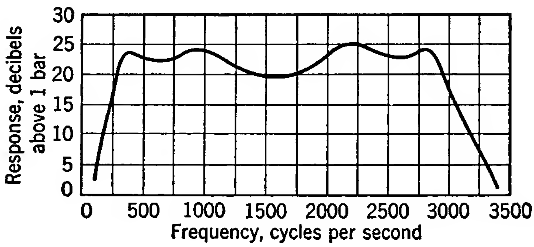


FIG. 37. Frequency response of the telephone receiver of Fig. 36. (Courtesy Automatic Electric Co.)

from 250 to 5000 cycles, has an output of about 110 decibels at 30 feet. Under ideal conditions this loudspeaker has been audible at 10 miles, and it is well suited to airports, harbors, freight yards, stadiums, and similar uses.

Frictional Loudspeakers. The Johnsen-Rahbek loudspeaker[5] consisted of a small, revolving cylinder of agate attached to one side of the incoming line. The other input wire was connected to a thin metal strip bearing upon the revolving cylinder and attached to a diaphragm. The friction between the cylinder and wire is caused to vary in accordance with the impressed electrical impulses, and thus sound signals are produced. A modification of this historical device using a revolving glass disk and a cork pad called a **Frenophone** has been developed.[38]

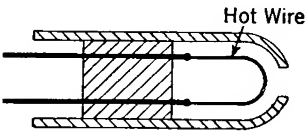


FIG. 38. A thermal receiver.

The Talking Arc. The talking or singing arc produces sound waves when electrical impulses are properly impressed across the arc.

Receivers operating on the principle of the condenser motor element (page 111) and on the principle of the crystal motor element (page 111) have been used. Also, high-quality moving-coil receivers[39] and ribbon receivers[40] have been developed for special purposes.

Acoustic Tests. It is difficult to make accurate tests of transmitters, microphones, receivers, and loudspeakers unless special apparatus and acoustically treated rooms are available. The transmitters and re-

ceivers used in telephone systems are usually tested with equipment designed for that specific purpose.[41] Microphones are usually compared with a standardized microphone in a room that is sound proof and acoustically treated so that reflections are negligible.

A telephone transmitter can be tested with some degree of reliability by placing the transmitter in a sound-proof box that is acoustically treated so that negligible reflection occurs from the inner walls. A small loudspeaker and the microphone of a sound-level meter (page 42) are also placed in the box. The microphone is placed adjacent to the transmitter under test and is used to measure the sound level as the sound frequency is varied, so that the level can be maintained constant. The transmitter under test is provided with normal direct current, and the signal output is measured with a vacuum-tube voltmeter, or by other suitable means.

A telephone receiver can be tested with a fair degree of reliability in a similar box. In this instance the receiver and the microphone of the sound-level meter are placed in the box, and the receiver is driven with an oscillator operating at suitable level. The sound-level meter gives an indication of the output at various frequencies. An oscilloscope and a wave analyzer also are useful in studying the performance of a transmitter or a receiver.

If a special acoustic laboratory is not available, the frequency response of a loudspeaker can be measured out-of-doors in a location free from reflecting objects. A simple method is to measure the intensity of the radiated sound with a sound-level meter. For such tests a beat-frequency oscillator provides an excellent source because the frequency is continuously variable and the loudspeaker can therefore be studied for any decided resonant points which often exist and might be missed if a point-to-point test were made. If measurements must be made in an untreated room, the reflected waves will cause points of maximum and minimum intensity to exist within the room; also, the location of these points will vary with frequency.

Power Output and Efficiency of Receivers and Loudspeakers. For the telephone receiver, if each of the values of Fig. 35 is multiplied by I^2 (the *effective value* of the testing current squared), a power diagram will be obtained. The value I^2Z_f is the *apparent power input;* the value I^2R_1 is the power dissipated in the ohmic resistance of the wire; I^2R_2 is the power dissipated electrically in skin-effect, eddy-current, and hysteresis losses, and I^2R_d is the total power delivered to the diaphragm. The various other terms may also be analyzed.[33]

Of the total effective power input to the diaphragm I^2R_d, little is radiated in the form of sound waves. For some telephone receivers,

the average ratio of the acoustic power output to the electrical power input is below 1 per cent.[42] Thus, if the input current of the receiver tested is 0.0025 ampere (a large input), the power input will be $I^2R = \overline{0.0025}^2 \times 210 = 0.00131$ watt, or 1310 microwatts. The value 210 ohms was obtained from Fig. 32 for a frequency of 820 cycles. Using an efficiency of 1 per cent, the power output in speech sounds would be 13.1 microwatts. This is a much larger output than under actual operating conditions. (Only about 10 microwatts of power is given out by the average voice in talking, page 30.)

A dynamic (moving-coil) loudspeaker will handle a continuous power input of about 5 watts. Auditorium dynamic loudspeakers will handle about 20 watts. Technical literature[43, 44] contains widely differing figures for the efficiency of dynamic loudspeakers; the variation is not surprising because of the many types available, the various uses, and the difficulty in measuring efficiencies. Based on references 43 and 44, and other sources of information, it is estimated that a dynamic loudspeaker in a baffle or cabinet has an efficiency of 2.5 to 10 per cent, that a dynamic loudspeaker in a directional baffle has an efficiency of 5 to 15 per cent, and that a dynamic loudspeaker in a short exponential horn has an efficiency of about 15 to 25 per cent. For the moving-coil dynamic unit driving a long exponential horn, the efficiency is estimated to be about 20 to 40 per cent. These units can handle a continuous power input of about 20 watts. It is emphasized that these are average estimated figures and that many exceptions exist. Standard methods have been devised[2] for determining the efficiency of loudspeakers.

In the design of sound systems for a theater[43] or an auditorium,[44] the *acoustic* power required must be known. The acoustic power requirements of auditoriums are given[44] as follows: 50,000 cubic feet, 1.2 watts; 100,000 cubic feet, 2.0 watts; 250,000 cubic feet, 4.0 watts; and 600,000 cubic feet, 8.0 watts. This same reference gives the equation

$$\text{Acoustic power (watts)} = 11.6\,\frac{VI}{T}, \tag{22}$$

where V is the volume of the auditorium in cubic feet, I is the sound intensity in watts per square centimeter, and T is the reverberation time in seconds. Also, this reference states that for orchestral reproduction the level should be +100 decibels (10^{-6} watt per square centimeter), based on the zero level of page 39, and that for speech the level should be +80 decibels (10^{-8} watt per square centimeter). These data assume average noise conditions.

REFERENCES

1. *American Standards of Electrical Terms.* A.I.E.E., 1941.
2. *Standards on Electroacoustics.* I.R.E., 1938.
3. *Report of the Standards Committee.* I.R.E., 1933.
4. Martin, W. H., and W. F. Davidson. The Trend in the Design of Telephone Transmitters and Receivers. *Bell System Tech. J.,* October, 1930, Vol. 9, No. 4.
5. Blake, G. G. *History of Radio Telegraph and Telephony.* Chapman & Hall, Ltd.
6. Thomas, P. A Diaphragmless Microphone for Radio Broadcasting. *J. A.I.E.E.,* March, 1923.
7. Shepardson, G. D. *Telephone Apparatus.* D. Appleton & Co.
8. Goucher, F. S. Microphonic Action in Telephone Transmitters. *Bell. Lab. Record,* August, 1930, Vol. 8, No. 12.
9. Goucher, F. S. The Carbon Microphone: An Account of Some Researches Bearing on Its Action. *Bell System Tech. J.,* April, 1934, Vol. 13, No. 2.
10. Christensen, C. J., and G. L. Pearson. Spontaneous Resistance Fluctuations in Carbon Microphones and Other Granular Resistances. *Bell System Tech. J.,* April, 1936, Vol. 15, No. 2.
11. Jones, W. C. Condenser and Carbon Microphones. *Bell System Tech. J.,* January, 1931, Vol. 10, No. 1.
12. Green, I. W., and J. P. Maxfield. Public Address Systems. *Trans. A.I.E.E.,* Vol. 42, 1923.
13. Wente, E. C. A Condenser Transmitter as a Uniformly Sensitive Instrument for Absolute Measurement of Sound Intensity. *Phys. Rev.,* July, 1917.
14. Wente, E. C. Electrostatic Transmitter. *Phys. Rev.,* May, 1922.
15. Olson, H. F., and F. Massa. *Applied Acoustics.* P. Blakiston's Son & Co.
16. Pamphlets of Western Electric Co. describing 640AA condenser microphone and RA-1095 amplifier.
17. Jones, W. C., and L. W. Giles. A Moving-Coil Microphone for High Quality Sound Reproduction. *Soc. Motion Picture Engrs.,* December, 1931, Vol. 17.
18. Wente, E. C., and A. L. Thuras. Moving-Coil Telephone Receivers and Microphones. *J. Acoust. Soc. Am.,* July, 1931.
19. Thuras, A. L. A Sensitive Moving-Coil Microphone of High Quality. *Bell Lab. Record,* May, 1932, Vol. 10, No. 9.
20. Marshall, R. N., and F. F. Romanow. A Non-Directional Microphone. *Bell System Tech. J.,* July, 1936, Vol. 15, No. 3.
21. Marshall, R. N. Low-Cost Microphone for Varied Application. *Bell Lab. Record,* November, 1937, Vol. 16, No. 3.
22. Marshall, R. N. New Cardioid Microphone. *Pick-Ups,* Western Electric Co., December, 1938.
23. Sivian, L. J. Absolute Calibration of Condenser Transmitters. *Bell System Tech. J.,* January, 1931, Vol. 10, No. 1.
24. Wente, E. C. Acoustical Instruments. *J. Acoust. Soc. Am.,* July, 1935, Vol. 7, No. 7.
25. Greaves, V. F., F. W. Kranz, and W. D. Crozier. The Kyle Condenser Loud Speaker. *Proc. I.R.E.,* July, 1929, Vol. 17, No. 7.
26. Williams, A. L. Piezo-Electric Loudspeakers and Microphones. *Electronics,* May, 1932, Vol. 4, No. 5.

27. Hewlett, C. W. A New Tone Generator. *Phys. Rev.*, Vol. 19, 1922.
28. Morse, A. H. *Radio: Beam and Broadcast.* D. Van Nostrand Co.
29. Hopkins, H. F. An Improved Loud-Speaking Telephone. *Bell. Lab. Record,* April, 1940, Vol. 18, No. 8.
30. Hanna, C. R., and J. Slepian. The Function and Design of Horns for Loud Speakers. *Trans. A.I.E.E.,* Vol. 43, 1924.
31. Hanna, C. R. Theory of Horn-Type Loud Speaker. *J. Acoust. Soc. Am.,* October, 1930.
32. Shepardson, G. D. Telephone Apparatus. D. Appleton & Co.
33. Kennelly, A. E. *Electrical Vibration Instruments.* The Macmillan Co.
34. Jones, W. C. Instruments for the New Telephone Sets. *Elec. Eng.,* October, 1938, Vol. 57, No. 10.
35. Pye, H. C. Transmission Characteristics of Types 40 and 50 Monophones. *Strowger Tech. J.* (Associated Electric Laboratories), December, 1939.
36. Rice, C. W., and E. W. Kellogg. Notes on the Development of a Hornless Loud Speaker. *Trans. A.I.E.E.,* Vol. 44, 1925.
37. Batcher, R. R. Sound Amplification by Air Modulation. *Electronic Ind.,* November, 1944, Vol. 3, No. 11.
38. Kershaw, J. B. C. Last Words of Science. *Sci. American,* January, 1924, Vol. 30.
39. Wolpert, F. S. A High-Quality Handset for Monitoring. *Bell Lab. Record,* July, 1937, Vol. 15, No. 11.
40. Olson, H. F., and F. Massa. A High-Quality Ribbon Receiver. *Proc. I.R.E.,* May, 1933, Vol. 21, No. 5.
41. Inglis, A. H., C. H. G. Gray, and R. T. Jenkins. A Voice and Ear for Telephone Measurements. *Bell System Tech. J.,* April, 1932, Vol. 12, No. 2.
42. Martin, W. H., and W. F. Davidson. The Trend in the Design of Telephone Transmitters and Receivers. *Bell System Tech. J.,* October, 1939, Vol. 9, No. 4.
43. Blattner, D. G., and L. G. Bostwick. Loud Speakers for Use in Theaters. *J. Acoust. Soc. Amer.,* October, 1930, Vol. 2.
44. Pender, H., and K. McIlwain. *Electrical Engineers' Handbook. Electric Communication and Electronics.* John Wiley & Sons, Inc.

REVIEW QUESTIONS

1. From a technical and a historical viewpoint, what is the distinction between the terms microphone and transmitter?
2. What is meant by the terms transducer, electroacoustic transducer, passive electroacoustic transducer, and active electroacoustic transducer?
3. What are the types of passive telephone transmitters? Of active transmitters? Which ones are used in practice?
4. How do the carbon granules function in a telephone transmitter?
5. What features are incorporated into handset transmitters so that they will be essentially non-resonant and non-positional?
6. What is meant by the term "capsule type" when applied to a telephone transmitter or receiver?
7. Distinguish between the use of the terms transmitter and receiver in telephony and radio.
8. For maximum power transfer, what should be the relation between the re-

sistance of a transmitter and the resistance of the circuit into which it is to work?

9. Discuss the nature and reasons for the distortion caused by telephone transmitters.
10. How does the so-called "sound-powered" telephone operate. Will it operate as both a transmitter and a receiver? Explain.
11. What is an important advantage of the double-button carbon microphone? An important disadvantage?
12. Discuss the condenser microphone. For what scientific purposes is it used?
13. What features do the moving-coil and ribbon microphones have in common? In what respects do they differ greatly?
14. Explain the operation of the crystal microphone. What important precaution must be taken in its use?
15. Define the terms telephone receiver and loudspeaker.
16. Name the two principal parts of a receiver or loudspeaker.
17. What types of electroacoustic motor elements are possible, and what types are in practical use?
18. Discuss the receivers used in telephone sets.
19. Why is a transformer commonly used with the radio dynamic loudspeaker?
20. What types of acoustic radiators are used with loudspeakers? Discuss the principle of operation of each.
21. For the horn-type loudspeaker, discuss the air chamber, the throat, the mouth, and the rate of taper.
22. Discuss the nature and reasons for the distortion caused by telephone receivers.
23. Discuss the nature of the input impedance of a telephone receiver. If any peculiarities exist, explain the causes.
24. What are some of the improvements incorporated in telephone receivers of the capsule type?
25. Briefly discuss some of the miscellaneous receivers and loudspeakers that have been used.
26. What important problems are encountered in making tests on electroacoustic devices?
27. What are the approximate average power inputs to a telephone receiver, a radio-set dynamic loudspeaker, a large horn-type loudspeaker?
28. Give the approximate efficiencies of each device in Question 27.
29. The length of a loudspeaker horn sometimes is quite objectionable. What is a remedy if this general type of radiator must be used? Explain diagrammatically.
30. What are several important factors that must be determined when a sound system is to be designed for an auditorium?

PROBLEMS

1. The resistance of a carbon-granule transmitter in the quiescent state is 70 ohms. When it is actuated by sound waves, the change in resistance is 12 per cent above and below this value. If a 3-volt battery and a 100-ohm resistor are connected in series with this transmitter, calculate the current that will flow and the power that will be delivered to the resistor.
2. Assume that the pressure on the diaphragm of the microphone of Fig. 9 is 10 bars and that the microphone is connected through an input transformer to

an amplifier. Calculate the approximate voltage amplification (both as a ratio and in decibels) that the amplifier must have to produce an output voltage of 25 volts.

3. The internal capacitance of a condenser microphone is assumed to be 50 micromicrofarads. Make the calculations required to show that the input resistor of the associated amplifier should be about 100 megohms, and that the amplifier should be close to the microphone with negligible connecting leads.
4. Use Fig. 11 as a basis, and at 500 cycles compute the dbm output at 0.1, 1.0, 10, and 100 bars.
5. Use the data accompanying Fig. 13, and calculate the open-circuit output voltage at 0.1, 1.0, 10, and 100 bars. If the output impedance is assumed to be 50 ohms resistance and 100 feet of microphone cable having a capacitance of 20 micromicrofarads per foot is used between the microphone and a 50-ohm resistance load, calculate the effect of the cable capacitance on the voltage delivered the load at 50, 1000, and 8000 cycles.
6. The output impedance of an amplifier is 15 ohms, and it is to be used to drive four loudspeakers with 15-ohm driving coils. How should the connections be made? If each loudspeaker draws 15 watts and if they are 350 feet from the amplifier, what size wire would you recommend for the line feeding the four speakers?
7. If the voice-coil current of the loudspeaker discussed on page 112 is 1.0 ampere, make an estimate of the power radiated, and of the efficiency. Does this compare with data given on page 132?
8. The diameter of the throat of a horn is 0.75 inch. If it is to transmit up to 7000 cycles, what should be the length of the horn and the diameter of the mouth, assumed to be circular.
9. Use the data for a telephone receiver given in Figs. 26 to 34 and draw to scale a diagram similar to Fig. 35. For the frequency of resonance, calculate the power lost in the direct-current resistance of the windings, the power lost in the magnetic circuit exclusive of the diaphragm, and the power delivered to the diaphragm. If this is assumed completely radiated, calculate the efficiency of the receiver. Compare this with the statement on page 132 and account for differences that may exist. Also compare with the calculations on page 132.
10. What acoustic power would you recommend for a college auditorium that will seat 5000 people? Assuming that you used horn-type speakers, what power-handling capacity should the amplifier have?

CHAPTER 5

ELECTRIC NETWORKS

An electric network is defined[1] as "a combination of any number of electric elements, the impedances of which may be either lumped or distributed or both, which are connected in any manner, conductively, inductively, or capacitively."

Electric networks may be classified[1] as **active electric networks** (a network containing one or more sources of energy) or as **passive electric networks** (a network containing no source of energy). It is understood that a network is passive and contains no source of energy *within itself* unless the contrary is stated.[1] Passive lumped networks only will be considered in this chapter.

Network Elements. Electric networks of lumped impedances are composed of elements such as resistors, capacitors, inductors and transformers. The characteristics of such circuit elements were treated in Chapter 3. In the pages that follow it will be assumed that the elements and the networks are **linear**; that is, have constants and transmission characteristics that do not vary with the magnitude of either the voltage impressed or the current flowing (page 85). Any circuit containing coils with ferromagnetic cores will be **non-linear,** but this effect is assumed negligible. The elements and networks are considered **bilateral,** passing current and signals equally well in both directions. Networks containing rectifying elements that are **unilateral** will not be considered.

Voltage Dividers. Communication circuits are often composed of a "chain" of devices and networks that passes electric signals from one device or network to the next. In many instances all the output voltage from one network is not needed by the next, or it may be advisable to make provisions for varying the voltage. **Voltage dividers** are used to select a fixed portion of an available voltage or to provide a variable voltage. These are often called potentiometers, a term erroneously used.[1]

An example of the use of a voltage divider is shown in Fig. 1. The voltage divider is a resistor with either fixed or movable contacts.[1] In communication the driving circuit often can supply but little energy, and the driven circuit requires but little energy. In such instances the voltage divider would be a high resistance of perhaps

500,000 ohms. In some instances a voltage divider must be of low resistance (such as for terminating a transmission line, page 202).

Electric Transducers. Many electric networks, and particularly those considered in this chapter, are **electric transducers,** defined[1] as "an electric network by means of which energy may flow from one or more transmission systems to one or more other transmission sys-

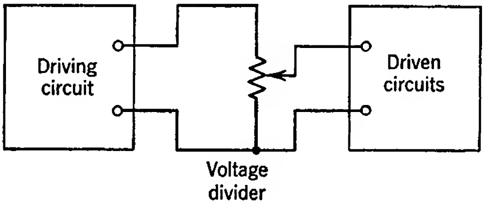


FIG. 1. A voltage divider is a resistor with fixed or movable contacts, often used as shown.

tems." An electric transducer generally has four terminals—two input terminals and two output terminals.[1] As for any electric network, electric transducers may be either *active* or *passive.* Those considered in this chapter are passive; that is, they contain within themselves no source of energy.

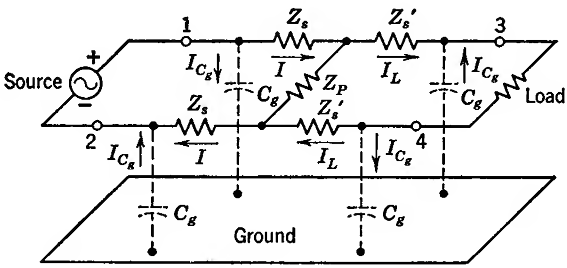


FIG. 2. A balanced network.

Balanced and Unbalanced Transducers. Most electric networks can be classified as two types on the basis of the arrangement of the impedance elements with respect to ground. The first type is the *balanced network* of Fig. 2. Assuming that the plane of the network is parallel to the ground plane, that the physical sizes of the Z_s elements are the same and that the physical sizes of the Z_s' elements are the same, the stray capacitances C_g to ground will be the same. If, also, the impedances of the series elements Z_s are identical, and the impedances Z_s' are identical, currents I will be the same and the currents I_L will be the

same. The series elements Z_s and Z_s' *on each side* of the circuit are paralleled by the stray capacitances to ground, and the current flowing through the load* is the sum of the currents I_L and I_{C_g} arriving by the two paths. If the identities previously discussed are maintained then the series voltage drops IZ_s will be the same; the series voltage drops I_LZ_s' will be the same; and the currents I_{C_g} will be the same. Then, the currents I in impedances Z_s will be identical, and the currents I_L in impedances Z_s' will be identical.

An *unbalanced network* is shown in Fig. 3; this is an extreme instance, the series impedances having been omitted in one side. The current that reaches Z_L will be I_L and I_{C_g} much as before, but the current path back to the source has been altered. Thus, I and I' are *not* equal, and I_L and I_L' are *not* equal. A circuit such as Fig. 2 (containing four series elements) will be unbalanced if *corresponding* series elements Z_s are not identical, if corresponding series elements Z_s' are not identical, or if the plane of the circuit is not parallel to the ground plane.

From the preceding discussion it follows that a **balanced network** is one in which the corresponding series impedance elements are identical, and one in which these elements are symmetrical (electrically) with respect to some reference (ground) potential. **Unbalanced networks** are those that do not fulfill these requirements.

The ground plane referred to may be the surface of the earth (as for an open-wire transmission line), may be the metal chassis of an amplifier, or may be a metal sheet under the top of a laboratory test bench. If it is necessary to "ground" a balanced circuit such as Fig. 2, the ground wire should be attached at the center of the parallel impedance Z_p; this maintains the balance to ground. An unbalanced circuit such as Fig. 3 is often grounded by attaching the ground wire to the side containing no series impedances.

Whether a circuit should be balanced or unbalanced and whether it should be grounded or ungrounded depend on circumstances. The important points at present are that fundamentally they are *different types of circuits* and that in general balanced and unbalanced circuits *must not be interconnected for testing or for operation, unless the connection is made through a transformer that has a grounded shield between the primary and the secondary.* If this is not done, measurements and operation at communication frequencies (including audio frequencies) will probably be unsatisfactory.

* By definition,[1] the word *load* means the power that an apparatus or machine delivers. Hence, strictly speaking the word "loading impedance" should be used in Fig. 2 and elsewhere in this book. It is accepted practice in communication to refer to the loading impedance as the load.

Because one side of the circuit contains essentially no impedance, it is possible to have one input and one output terminal common. Thus, Fig. 3 is sometimes referred to as a three-terminal network, and Fig. 2 is sometimes referred to as a four-terminal network. This classification is of little (if any) fundamental importance, *but whether or not a circuit is balanced or unbalanced is of great fundamental importance.*

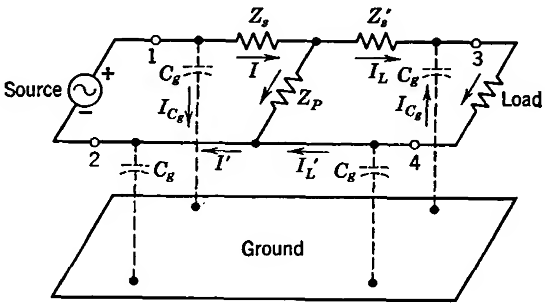


FIG. 3. An unbalanced network.

Symmetrical and Unsymmetrical Transducers. In discussing the network of Fig. 2 it was pointed out that for balance the two Z_s elements must be identical, and the two Z_s' elements must be identical. For this network to be both balanced and symmetrical, *each* of the series elements must be identical. If this is true, then the network of Fig. 2 offers the same input impedance characteristics when considered from terminals 1–2 or terminals 3–4.

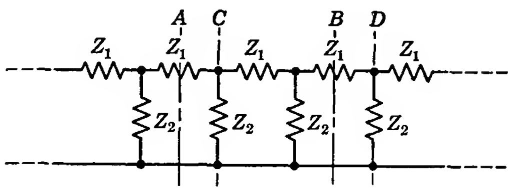


FIG. 4. A portion of an unbalanced network.

A network may be unbalanced, yet symmetrical. Thus in Fig. 3, if Z_s is of the same impedance as Z_s', then the unbalanced circuit will be symmetrical. The network (only) will offer the same input impedance characteristics when considered from terminals 1–2 or terminals 3–4 (see also page 154).

Midseries and Midshunt Terminations. Often several networks are connected one after the other in a "chain," in "cascade," or in

"tandem." In such a transmission system, as far as the signal is concerned, the individual networks tend to lose their identity. It is advisable in network study first to investigate the system as a whole.

Figure 4 shows an unbalanced network extending to infinity in both directions. It is desired to select a finite portion of this network to use as a transmission system. Suppose the portion A–B is selected by cutting the network at $\frac{1}{2}Z_1$. This will give the network of Fig. 5*a*, and the network is said to have **mid-series terminations,** because it "starts" and "ends" in $\frac{1}{2}Z_1$.

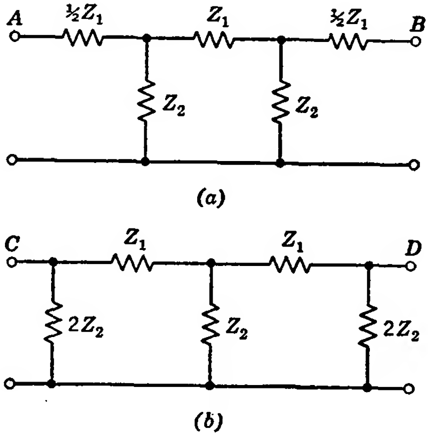


FIG. 5. The circuit of (*a*) is obtained by selecting the portion A–B of Fig. 4. The circuit of (*b*) is obtained by selecting the portion C–D of Fig. 4.

Another possibility is to select the finite section C–D (Fig. 5*b*). This network is said to have **mid-shunt terminations.** The value $2Z_2$ is specified because two such impedances in parallel give the impedance Z_2 of Fig. 4. The infinite network *could* have been cut into at any point; then fractional termination would have resulted.

T and π Sections. Assume that it is desired to construct a network such as Fig. 5(*a*). Because two series impedances $\frac{1}{2}Z_1$ add to give an impedance Z_1, the circuit of Fig. 5(*a*) can be formed of the two **T sections** of Fig. 6(*a*). Also, a network such as Fig. 5(*b*) can be formed of the two **π sections** of Fig. 6(*b*).

Equivalence of T and π Sections. In communication circuits it sometimes is desired to find the π section that is equivalent to a T section, or vice versa. Two networks are of **general equivalence**[1] "if one network can replace another network in any system whatsoever without altering in any way the electrical operation of that portion of the system external to the network." Two networks are of **limited equivalence**[1] "if one network can replace another network only in some particular system without altering in any way the electrical operation of that portion of the system external to the networks." In the following discussion *limited equivalence* will be investigated. It is very important to note that networks often are constructed of

impedances that vary with frequency; usually, *networks are equivalent only at a given frequency.*

In Fig. 7 are shown T and π networks. To derive the **transformation equations** from which an equivalent π section can be designed if

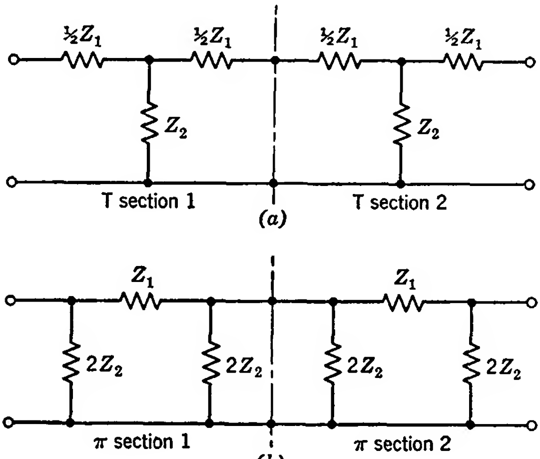


Fig. 6. T and π sections.

a T section is known (or vice versa) three equations will be written.

For the T section, the impedance Z_{12} measured between terminals

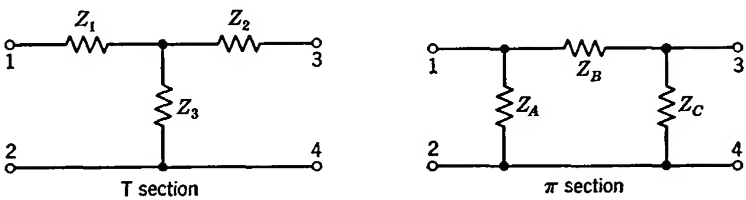


Fig. 7. Circuits for deriving network transformation equations.

1–2 is $Z_{12} = Z_1 + Z_3$, and for the π section the impedance $Z_{12} = Z_A(Z_B + Z_C)/(Z_A + Z_B + Z_C)$. When these are equated,

$$Z_{12} = Z_1 + Z_3 = \frac{Z_A(Z_B + Z_C)}{Z_A + Z_B + Z_C}. \tag{1}$$

The corresponding equations for impedances measured between terminals 3–4 are

$$Z_{34} = Z_2 + Z_3 = \frac{Z_C(Z_A + Z_B)}{Z_A + Z_B + Z_C}. \tag{2}$$

The corresponding equations for impedances measured between terminals 1–3 are

$$Z_{13} = Z_1 + Z_2 = \frac{Z_B(Z_A + Z_C)}{Z_A + Z_B + Z_C}. \tag{3}$$

These three equations can be solved simultaneously for Z_1, Z_2, and Z_3 in terms of Z_A, Z_B, and Z_C. By adding equations 1 and 3, and subtracting equation 2, it is found that

$$Z_1 = \frac{Z_A Z_B}{Z_A + Z_B + Z_C} \tag{4}$$

Likewise, adding equations 2 and 3, and subtracting equation 1 gives

$$Z_2 = \frac{Z_B Z_C}{Z_A + Z_B + Z_C}, \tag{5}$$

and adding equations 1 and 2, and subtracting equation 3 gives

$$Z_3 = \frac{Z_A Z_C}{Z_A + Z_B + Z_C}. \tag{6}$$

These equations are for transformation from a π to a T network. To make transformations from T to a π network, the equations prove to be[2, 3]

$$Z_A = \frac{Z_1Z_2 + Z_2Z_3 + Z_1Z_3}{Z_2}, \tag{7}$$

$$Z_B = \frac{Z_1Z_2 + Z_2Z_3 + Z_1Z_3}{Z_3}, \tag{8}$$

and

$$Z_C = \frac{Z_1Z_2 + Z_2Z_3 + Z_1Z_3}{Z_1} \tag{9}$$

It is now desired to show that if the two networks are related according to these equations they are equivalent and can replace each other without altering the circuit operation. For this Fig. 8 will be used, in which the two circuits, *related in accordance with the equations just given,* are connected between *identical* sources of voltage E_g, and *identical* load impedances Z_L. If it can be shown that in each circuit the identical generators supply currents I_1 that are exactly the same and that the identical load impedances Z_L receive currents I_2 that are exactly the same, then the T and π sections are equivalent as far as terminals 1–2 and 3–4 are concerned.

For the T section of Fig. 8,

$$I_1 = \frac{E_g}{Z_g + Z_1 + \dfrac{(Z_2 + Z_L)Z_3}{Z_2 + Z_3 + Z_L}}. \tag{10}$$

Because the voltage drops across two parallel branches must be equal,

$$(I_1 - I_2)Z_3 = I_2(Z_2 + Z_L), \quad \text{and} \quad I_2 = \frac{I_1 Z_3}{Z_2 + Z_3 + Z_L}. \tag{11}$$

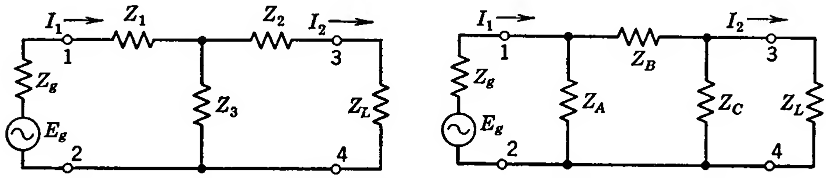


FIG. 8. Networks for studying conditions of limited equivalence.

The next step is to substitute in these two equations the values of Z_1, Z_2, and Z_3 given in equations 4, 5, and 6. Making these substitutions in equation 10 and simplifying the expression gives

$$I_1 = \frac{E_g}{Z_g + Z_A\left[\dfrac{Z_L(Z_B + Z_C) + Z_B Z_C}{(Z_A + Z_B)(Z_L + Z_C) + Z_L Z_C}\right]} \tag{12}$$

Making these substitutions in equation 11 gives

$$I_2 = \frac{I_1 Z_A Z_C}{(Z_A + Z_B)(Z_L + Z_C) + Z_L Z_C}. \tag{13}$$

It now remains to derive expressions for I_1 and I_2 for the π section of Fig. 8, in terms of Z_A, Z_B, and Z_C and to ascertain if these expressions agree with equations 12 and 13. Thus, for the π section of Fig. 8,

$$I_1 = \frac{E_g}{Z_g + \dfrac{Z_A\left(Z_B + \dfrac{Z_C Z_L}{Z_C + Z_L}\right)}{Z_A + Z_B + \dfrac{Z_C Z_L}{Z_C + Z_L}}},$$

and simplifying,

$$I_1 = \frac{E_g}{Z_g + Z_A\left[\dfrac{Z_L(Z_B + Z_C) + Z_B Z_C}{(Z_A + Z_B)(Z_L + Z_C) + Z_L Z_C}\right]} \tag{14}$$

To find the current I_2 in the π section of Fig. 8 first it is necessary to find the current in impedance Z_B, which can be determined from the relation

$$(I_1 - I_{ZB})Z_A = I_{ZB}\left(Z_B + \frac{Z_C Z_L}{Z_C + Z_L}\right)$$

and

$$I_{ZB} = \frac{I_1 Z_A}{Z_A + Z_B + \dfrac{Z_C Z_L}{Z_C + Z_L}}$$

Similarly, $(I_{ZB} - I_2)Z_C = I_2 Z_L$, and $I_2 = \dfrac{I_{ZB} Z_C}{Z_L + Z_C}$.

Substituting the value of I_{ZB} in this expression gives

$$I_2 = \frac{I_1 Z_A Z_C}{(Z_A + Z_B)(Z_L + Z_C) + Z_L Z_C}. \tag{15}$$

It has been shown that equations 12 and 13 derived for the T section are identical to equations 14 and 15 derived for the π section. This

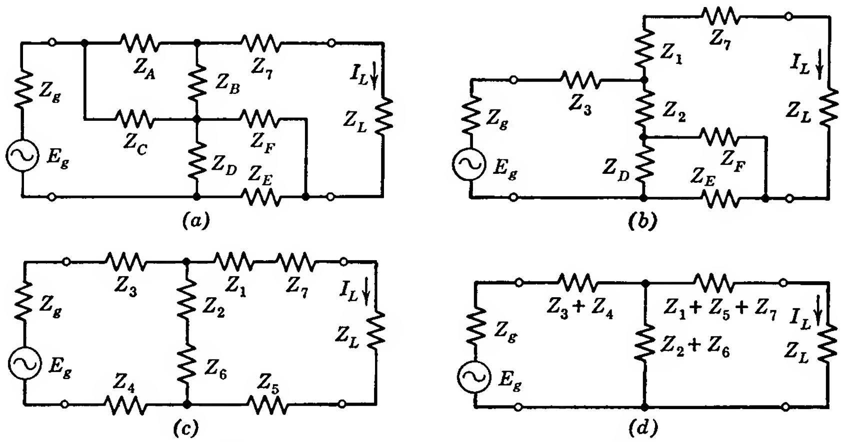


FIG. 9. Diagrams indicating the steps in simplifying circuits.

proves that the sections will perform exactly alike when inserted between a source of voltage and a load impedance and that *so far as terminals 1–2 and 3–4 are concerned* T and π networks related by equations 4, 5, and 6, or 7, 8, and 9 are equivalent.

Simplification of Networks. Communication networks are often complicated, and it sometimes is advantageous to reduce them to simple networks before a final solution is made. Thus, in Fig. 9(a) suppose

that it is desired to find the current I_L that the generator of voltage E_g and of internal impedance Z_g sends through the load impedance Z_L.

The combination Z_A, Z_B, and Z_C can be treated as a π section, and the equivalent T section can be obtained from equations 4, 5, and 6. This gives the circuit of Fig. 9(*b*). These new impedances are called Z_1, Z_2, and Z_3.

The combination Z_D, Z_E, and Z_F also can be treated as a π section, and the equivalent T section can be found from equations 4, 5, and 6, giving the circuit of Fig. 9(*c*) with the new impedances called Z_4, Z_5, and Z_6.

All series impedances are then added, giving Fig. 9(*d*), which can readily be solved for the desired current I_L through the load impedance Z_L.

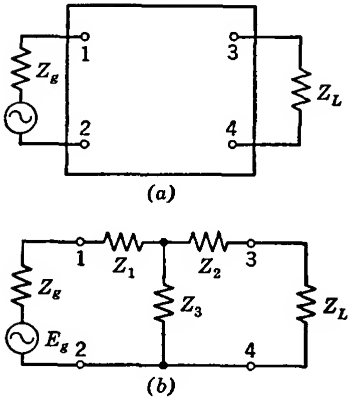


FIG. 10. Unknown network (*a*), and equivalent network (*b*).

Equivalent Networks. In the preceding section the value of each impedance of Fig. 9 (*a*) was known, and the reduction to the simple T network was made by transformation equations. Sometimes a circuit or device is in a "black box" (Fig. 10), the values of the impedances are unknown, and it is desired to determine the simple equivalent T or π network.

For limited equivalence between the two sets of input terminals 1–2 and the two sets of output terminals 3–4, the input currents I_1 to the two networks must be identical and the output currents I_2 to the loads Z_L must be identical (in both magnitude and phase). Of course, this also implies that the voltage across terminals 1–2 must be identical, and the output voltages across terminals 3–4 must be identical when the networks are connected between the same generator and the same load impedance Z_L.

An investigation will show that if the input current I_1 does *not* equal the output current I_2, there must be a shunt element such as Z_3 in an equivalent network (Fig. 9 or 10). Similarly, if the input voltage E_{12} does *not* equal the output voltage E_{34}, there must be series elements such as Z_1 and Z_2. Thus, *three* impedances are required in an equivalent network. These three impedances may be in a T configuration as in Fig. 10, or may be transformed into a π configuration.

The individual elements within the black box of Fig. 10 in general cannot be determined. It is known, however, that, at a given frequency, some T network will be equivalent. The values of Z_1, Z_2,

and Z_3 of this equivalent T network must be determined from measurements on the unknown network. Thus, if the input impedance between terminals 1–2 is measured with the output terminals 3–4 open circuited and if this is called Z_{12oc}, then

$$Z_{12oc} = Z_1 + Z_3. \tag{16}$$

If the input impedance Z_{12sc} is measured between input terminals 1–2 with the output terminals 3–4 short circuited, then

$$Z_{12sc} = Z_1 + \frac{Z_2 Z_3}{Z_2 + Z_3}. \tag{17}$$

If the output impedance Z_{34oc} is measured between output terminals 3–4 with the input terminals 1–2 open circuited, then

$$Z_{34oc} = Z_2 + Z_3 . \tag{18}$$

From these three equations,

$$Z_1 = Z_{12oc} - Z_3 , \tag{19}$$

$$Z_2 = Z_{34oc} - Z_3, \tag{20}$$

and

$$Z_3 = \sqrt{(Z_{12oc} - Z_{12sc})Z_{34oc}} . \tag{21}$$

Equation 21 is found by substituting equations 19 and 20 in equation 17 and solving for Z_3.

The Lattice Network. The networks (such as Fig. 4) considered in the preceding pages are known as **ladder networks** because of the arrangement of the series and shunt elements. Another fundamental type is the **lattice network** of Fig. 11. Considering only the input terminals 1–2 and the output terminals 3-4, a T (or a π) ladder network, such as Fig. 10(b), exists that is equivalent to the lattice network, and this limited equivalent network can be found from the three impedance measurements, Z_{12oc}, Z_{12sc}, Z_{34oc}.

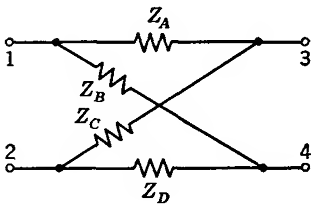


Fig. 11. A lattice network.

From Fig. 11 it is seen that

$$Z_{12oc} = \frac{(Z_A + Z_C)(Z_B + Z_D)}{Z_A + Z_B + Z_C + Z_D}, \tag{22}$$

$$Z_{12sc} = \frac{Z_A Z_B}{Z_A + Z_B} + \frac{Z_C Z_D}{Z_C + Z_D}, \tag{23}$$

and that

$$Z_{34oc} = \frac{(Z_A + Z_B)(Z_C + Z_D)}{Z_A + Z_B + Z_C + Z_D}. \tag{24}$$

When these terms are inserted in equations 21, 20, and 19 in this order and when equations are simplified,[3]

$$Z_3 = \frac{Z_B Z_C - Z_A Z_D}{Z_A + Z_B + Z_C + Z_D}, \tag{25}$$

$$Z_2 = \frac{Z_A Z_C + 2Z_A Z_D + Z_B Z_D}{Z_A + Z_B + Z_C + Z_D}, \tag{26}$$

and

$$Z_1 = \frac{Z_A Z_B + 2Z_A Z_D + Z_C Z_D}{Z_A + Z_B + Z_C + Z_D}. \tag{27}$$

Network Theorems. The theorems now to be discussed are useful in studying networks. It is assumed that the networks are composed of elements that are linear and bilateral, that the steady-state condition has been reached, and that all quantities are complex.

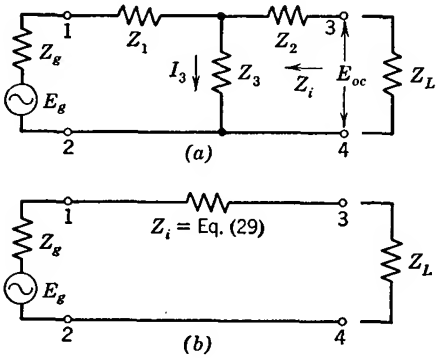


FIG. 12. Circuits for studying Thévenin's theorem.

Thévenin's Theorem. If an impedance Z_L is connected between any two points of a circuit, the current I_L that will flow through this impedance is the same as if the impedance Z_L is connected to a generator whose constant generated voltage E_g is the same as the voltage E_{oc} between the two points of the circuit, prior to the connection, and whose internal impedance Z_g is the same as the impedance Z_i measured, prior to the connection, back into the circuit, with each source of driving voltage within the circuit replaced with an impedance equal in value to the internal impedance of that source.

To prove Thévenin's theorem, recall that any passive network can be represented by the T section of Fig. 12. If a generator is connected as shown, an open-circuit voltage will appear between terminals 3–4. The magnitude of this open-circuit voltage is

$$E_{oc} = I_3 Z_3 = \frac{E_g Z_3}{Z_g + Z_1 + Z_3}. \tag{28}$$

The impedance measured toward the generator from terminals 3–4 is

$$Z_i = Z_2 + \frac{Z_3(Z_1 + Z_g)}{Z_g + Z_1 + Z_3}. \tag{29}$$

According to Thévenin's theorem, when Z_L is connected

$$I_L = \frac{E_{oc}}{Z_i + Z_L} = \frac{\text{Equation 28}}{\text{Equation 29} + Z_L} =$$

$$\frac{E_gZ_3}{Z_2Z_g + Z_1Z_2 + Z_2Z_3 + Z_1Z_3 + Z_3Z_g + Z_gZ_L + Z_1Z_L + Z_3Z_L}. \tag{30}$$

From ordinary circuit theory, when Z_L is connected,

$$I_g = \frac{E_g}{Z_g + Z_1 + \dfrac{Z_3(Z_2 + Z_L)}{Z_2 + Z_3 + Z_L}}. \tag{31}$$

Since the voltage drops across Z_3 and $(Z_2 + Z_L)$ of Fig. 12 must be equal,

$$I_L(Z_2 + Z_L) = (I_g - I_L)Z_3, \quad \text{and} \quad I_L = \frac{I_gZ_3}{Z_2 + Z_3 + Z_L}. \tag{32}$$

Substituting equation 31 in equation 32 gives

$$I_L = \frac{E_gZ_3}{Z_2Z_g + Z_1Z_2 + Z_2Z_3 + Z_1Z_3 + Z_3Z_g + Z_gZ_L + Z_1Z_L + Z_3Z_L}. \tag{33}$$

Thus, since equations 30 and 33 are identical, the proof of Thévenin's theorem is established. Regarding the last portion of Thévenin's theorem, consult the superposition theorem in the following pages.

Norton's Theorem. If an impedance Z_L is connected between any two points of a circuit, the current I_L that will flow through this impedance is the same as if the impedance Z_L is connected to a generator whose constant generated current I_g is the same as the current I_{sc} that flows if the two points of the circuit are short circuited, the constant-current generator being in parallel with an impedance Z_i measured prior to the connection back into the circuit, with each other source of driving voltage within the circuit replaced with an impedance equal in value to the internal impedance of that source.

In proving Norton's theorem, it will be recalled that any circuit can be reduced to the T network of Fig. 13(*a*) and it must be shown that this circuit is equivalent to Fig. 13(*b*). Since Fig. 13(*a*) already has

been shown to be equal to Fig. 12(*b*), if an equivalence between Fig. 12(*b*) (also reproduced as Fig. 13(*c*)) and Fig. 13(*b*) is established, then Norton's theorem will be proved.

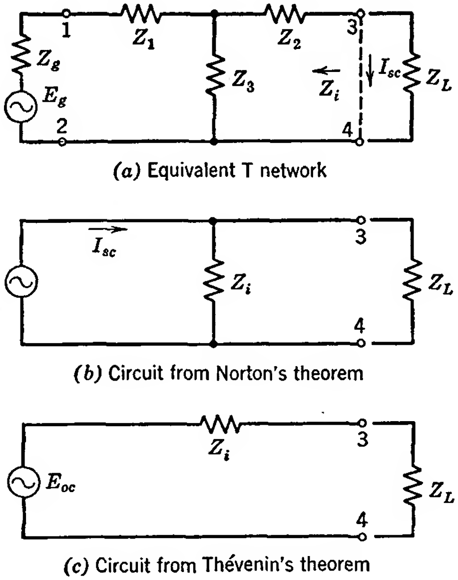


FIG. 13. The circuit of (*a*) may be solved by the use of either Norton's theorem or Thévenin's theorem.

From Thévenin's theorem, the current that will flow through a load impedance Z_L is

$$I_L = \frac{E_{oc}}{Z_i + Z_L}, \tag{34}$$

and, if the generator of Fig. 13(*c*) is short-circuited,

$$I_{sc} = \frac{E_{oc}}{Z_i} \quad \text{and} \quad E_{oc} = I_{sc}Z_i\,. \tag{35}$$

According to Norton's theorem and the circuit of Fig. 13(*b*),

$$(I_{sc} - I_L)Z_i = I_LZ_L \quad \text{and} \quad I_L = \frac{I_{sc}Z_i}{Z_i + Z_L}. \tag{36}$$

From equation 35, equation 36 becomes

$$I_L = \frac{E_{oc}}{Z_i + Z_L}, \tag{37}$$

which is the same as equation 34, thus establishing the fact that Norton's theorem and the circuit of Fig 13(*b*) can be used in solving network problems.

Norton's theorem is convenient for use with circuits of high impedance, such as voltage amplifiers employing tetrodes and pentodes having internal impedances (plate resistances) of about 1,000,000 ohms. Thévenin's theorem is convenient for low-impedance circuits.

Superposition Theorem. If a network has more than one source of driving voltage, the current that flows at any point, or the voltage between any two points, is the sum of the currents or voltages at these points which would exist if each source of voltage were considered separately, each of the other sources being replaced at that time by an impedance equal to the internal impedance of that source.

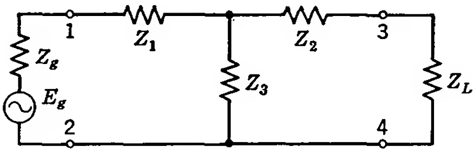


FIG. 14. Circuit for studying the reciprocity theorem.

If the impedances in a network are linear, and their values are not affected by the magnitudes of the currents and voltages existing, then the current or voltage in the various portions of the circuit are independent of each other and the principle of superposition applies. A simple laboratory experiment will verify this principle.

This theorem can be applied to a circuit, and the use of the Kirchhoff's law method avoided. Also, the effect of a given generator voltage (of the same or different frequency) at a given point can be determined without considering the effects of other voltage sources.

Reciprocity Theorem. If any source of voltage E located at one point in a network produces a current I at any other point in the network, the same source of voltage E placed at the second point will produce the same current I at the first point.

Any complex network can be reduced to the simple T section of Fig. 14. The current I_L that generator E_g forces through impedance Z_L is given by equation 33. If the *source* of voltage E_g *only* is removed and placed in series with Z_L and if equations such as 31, 32, and 33 for the current I_g through Z_g are written, it will be found that the two currents are identical.

Compensation Theorem. Any impedance in an energized network may be replaced by a generator of zero internal impedance whose instantaneous generated voltage is equal to the instantaneous potential difference across the replaced impedance caused by the current flowing through the impedance.

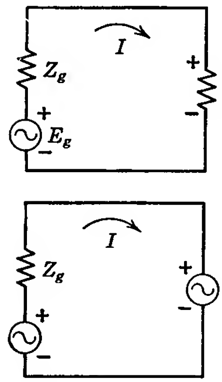


FIG. 15. Circuits for studying the compensation theorem.

The principle on which this theorem is based is well known. Thus in Fig. 15, if the polarity of the instantaneous generated voltage used to replace the impedance as indicated and if the magnitude of this generated voltage equals the magnitude of the potential difference across the impedance, the currents in the two circuits will be identical.

Maximum Power Transfer Theorems. If a passive network is joined to an active network by two terminals the power that will be absorbed by the passive network will be maximum if the impedances measured in the two directions, prior to the connection, are conjugates.

This was considered on page 70, and additional discussion is unnecessary, except to point out that, since any network can be replaced by a simple T structure and because Thévenin's theorem applies, this theorem can be represented by Fig. 16. For maximum power transfer, if $Z_g = R + jX$, then Z_L should be $Z_L = R - jX$, and vice versa. **Conjugate impedances** have equal resistances and equal but opposite reactances.

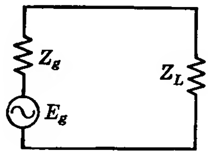


FIG. 16. Circuit for determining maximum power transfer relations.

If the magnitude, but not the angle, of a load impedance may be varied, then the maximum power will be absorbed when the magnitude of the load impedance equals the magnitude of the internal impedance of the source.

This was considered on page 71, and additional discussion is unnecessary.

Transformer Equivalent Networks. For the purposes of circuit analysis, it is sometimes desired to replace a transformer by an equivalent T (or π) network. The equivalent T network for a transformer can be found from three impedance measurements and the use of equations 19, 20, and 21. From Fig. 17 it is seen that $Z_{12_{oc}} = Z_p$, the impedance of the primary of the transformer; also, that $Z_{34_{oc}} = Z_s$, the impedance of the secondary. The impedance of the primary, with the secondary short circuited is, from coupled-circuit theory (page 67),

$Z_{12sc} = Z_p + (\omega M)^2/Z_s$. Thus, it is possible from these three measurements to determine both the important constants of a transformer, and the equivalent T network.

If the values just discussed are inserted in equations 21, 20, and 19 in the order given,

$$Z_3 = \sqrt{Z_p - (Z_p + \omega^2 M^2/Z_s)Z_s} = \omega M, \tag{38}$$

$$Z_2 = Z_s - \omega M = Z_s - Z_m, \tag{39}$$

and

$$Z_1 = Z_p - \omega M = Z_p - Z_m. \tag{40}$$

The reason that $\omega M = Z_m$, where Z_m is called the **mutual impedance,** can be explained in the following manner. From Fig. 17, when the secondary is open, the voltage E_{34} across the secondary is equal to the I_pZ_3 voltage drop. Thus, $Z_3 = E_{34}/I_p = \omega MI_p/I_p = \omega M = Z_m$, the term mutual impedance being used because, for the actual transformer of Fig. 17, the *secondary* voltage E_{34} is being divided by the *primary* current I_p.

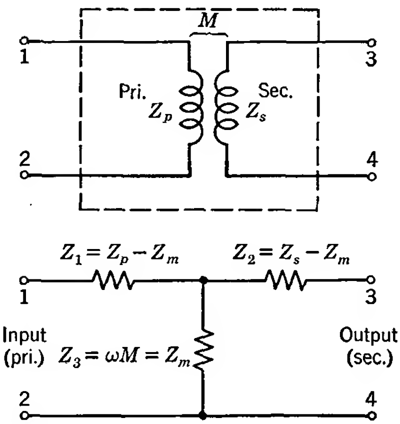


FIG. 17. Actual and equivalent circuits for a transformer.

Transformers are used in communication circuits for impedance matching (page 68). For such purposes, an **ideal transformer** is assumed. Such a transformer will change the magnitude of the load impedance without altering the angle of the load impedance and introduces no losses into the circuit. An ideal transformer is assumed in making possible maximum power transfer under the conditions of the second maximum power transfer theorem.

Image Impedances. Four-terminal transducers are used for passing electric energy from one part of an electric system to another. Often it is desired that the transducers draw maximum possible signal power from the preceding part of the system and pass it on to the following part of the system. Ordinarily, this cannot be done under the condition of conjugate impedances, and hence the conditions of the second power transfer theorem must be met.

The **image impedances** of a transducer are defined[1] as the "impedances which will simultaneously terminate each pair of terminals

of a transducer in such a way that at each pair of terminals the impedances in both directions are equal."

An unsymmetrical, unbalanced (page 139) T network is shown in Fig. 18, the network being unsymmetrical because Z_1 *does not* equal Z_2. It is desired to find the two image impedances Z_I and Z_I' that

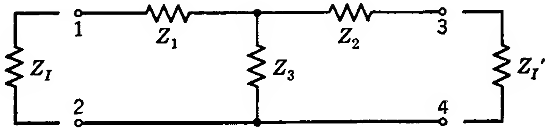


FIG. 18. Circuit for studying image impedances.

will satisfy the definition previously given. If this is accomplished, then maximum power transfer would occur from the transducer to the impedance (or from the impedance to the transducer) at either set of terminals.

The equations for the image impedances can be written for the conditions specified in the definition. Thus, the input impedance at terminals 1–2, with image impedance Z_I' connected at terminals 3–4 (but with Z_I *not* connected), is

$$Z_I = Z_1 + \frac{Z_3(Z_2 + Z_I')}{Z_2 + Z_3 + Z_I'}. \tag{41}$$

The input impedance at terminals 3–4, with image impedance Z_I connected at terminals 1–2 (but with Z_I' not connected), is

$$Z_I' = Z_2 + \frac{Z_3(Z_1 + Z_I)}{Z_1 + Z_3 + Z_I}. \tag{42}$$

When these two equations are solved, the image impedances are found to be

$$Z_I = \sqrt{\frac{Z_1 + Z_3}{Z_2 + Z_3}(Z_1Z_2 + Z_2Z_3 + Z_1Z_3)}, \tag{43}$$

and

$$Z_I' = \sqrt{\frac{Z_2 + Z_3}{Z_1 + Z_3}(Z_1Z_2 + Z_2Z_3 + Z_1Z_3)}. \tag{44}$$

It is also possible to write equations for the image impedances of a π section; or the π section can be transformed into a T section, and equations 43 and 44 then used. An investigation will disclose that

unsymmetrical networks have impedance-transforming properties and that they can be used for impedance matching.

Iterative Impedance. The circuit of Fig. 18, for which the image impedances were found, was unsymmetrical, and the two terminating image impedances were different. If a circuit is symmetrical, however, the image impedances become the same, and equal the iterative impedances[1] now to be discussed.

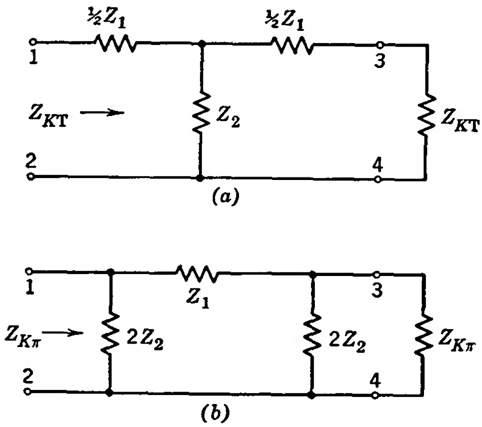


FIG. 19. Circuits for deriving the equations for midseries and midshunt iterative impedances.

The **iterative impedance** at a pair of terminals such as 1–2 or 3–4 of Fig. 19 is defined[1] as "the impedance which will terminate the other pair of terminals in such a way that the impedance measured at the first pair of terminals is equal to this terminating impedance." For an unsymmetrical section the iterative impedances at the two sets of terminals are different; for a symmetrical section the iterative impedances are equal and the same as the image impedances.[1] Because of their wide use, symmetrical sections will be treated in the remainder of this section.

The notation used for the elements of the T and π sections of Fig. 19 are the same as was used for the sections of Fig. 6. This notation is usually employed for the circuits now to be considered.

The input impedances at terminals 1–2 of Fig. 19(a) when terminals 3–4 are terminated in the iterative impedance Z_{KT} of the T network is

$$Z_{KT} = \frac{1}{2}Z_1 + \frac{Z_2(\frac{1}{2}Z_1 + Z_{KT})}{\frac{1}{2}Z_1 + Z_2 + Z_{KT}},$$

which, when solved for Z_{KT}, gives

$$Z_{KT} = \sqrt{Z_1Z_2 + \frac{Z_1^2}{4}} = \sqrt{Z_1Z_2\left(1 + \frac{Z_1}{4Z_2}\right)}. \tag{45}$$

For the π section, the iterative impedance $Z_{K\pi}$ is the input impedance at terminals 1–2 of Fig. 19(b) when terminals 3–4 are terminated in the iterative impedance $Z_{K\pi}$ of the π network. Thus

$$Z_{K\pi} = \frac{2Z_2\left(Z_1 + \dfrac{2Z_2Z_{K\pi}}{2Z_2 + Z_{K\pi}}\right)}{2Z_2 + Z_1 + \dfrac{2Z_2Z_{K\pi}}{2Z_2 + Z_{K\pi}}} = \frac{Z_1Z_2}{\sqrt{Z_1Z_2 + \dfrac{Z_1^2}{4}}} = \frac{Z_1Z_2}{Z_{KT}}. \tag{46}$$

The iterative impedance of a network such as shown in Fig. 19(a) or (b) is readily determined from open-circuit and short-circuit impedance measurements. Thus, if an impedance bridge is used to measure the input impedance at terminals 1–2 of the T section of Fig. 19(a) when terminals 3–4 are open,

$$Z_{12\text{oc}} = \tfrac{1}{2}Z_1 + Z_2, \tag{47}$$

and the input impedance at terminals 1–2 with terminals 3–4 short circuited is

$$Z_{12\text{sc}} = \frac{1}{2}Z_1 + \frac{\frac{1}{2}Z_1Z_2}{\frac{1}{2}Z_1 + Z_2}. \tag{48}$$

If these two equations are multiplied together and the square root taken,

$$\sqrt{Z_{\text{oc}}Z_{\text{sc}}} = \sqrt{\left(\frac{1}{2}Z_1 + Z_2\right)\left(\frac{1}{2}Z_1 + \frac{\frac{1}{2}Z_1Z_2}{\frac{1}{2}Z_1 + Z_2}\right)} = \sqrt{Z_1Z_2 + \frac{Z_1^2}{4}} = Z_{KT}. \tag{49}$$

Although this derivation was for a T section, it can be shown that the iterative impedance of a π section also can be found from open-circuit and short-circuited measurements.

Midseries and Midshunt Iterative Impedance. Suppose that another identical T section is connected ahead of terminals 1–2 of Fig. 19(a). The input impedance to this new T section also will be Z_{KT}, because, in accordance with the definition for iterative impedance, this new T section also will be terminated with Z_{KT}. The same reasoning applies, of course, to the π section of Fig. 19(b).

Now suppose that an infinite number of T sections (or an infinite

number of π sections) are in tandem forming an iterative or recurrent structure. The input impedance of the infinite number of T sections, or the input impedance of the infinite number of π sections, also will be the iterative impedance, whether or not the distant end at infinity is terminated in the iterative impedance. The line is so long that an electric signal never reaches the distant end, and hence it makes no difference how the end at infinity is terminated. Thus, the input impedance of an infinite number of identical recurrent sections equals the iterative impedance of a network.

As was shown in Figs. 4, 5, and 6, a group of T sections or π sections, when connected one after the other, "blend into" the same basic structure (Fig. 4). The only difference between the two will be at the sending and receiving ends. Here the end sections will be "one half" of a T section or "one half" of a π section, giving a midseries termination or a midshunt termination (page 141). Thus, except from the standpoint of the performance at the input or the output terminals, it makes no difference whether a recurrent group of identical networks is T sections, π sections, or "blended together" as in Fig. 4. The matter of importance is how the structure is terminated. If it is terminated in midseries, then the input (and ouput) impedance is the **midseries iterative impedance** given by equation 45; if it is terminated in midshunt, then the input (and output) impedance is the **midshunt iterative impedance** given by equation 46.

Transition Loss. As has been mentioned elsewhere, electric transducers, such as the networks considered in this chapter, must pass electric signal energy from one network to the next, etc. As shown on pages 71 and 152, the ability to pass energy from one network to the next depends on the impedance relations at the network junctions. If the output impedance of one network is not related correctly to the input impedance of the next network, then conditions for maximum power transfer will not exist, and a **transmission loss** or **loss,** defined[1] as "a term used to denote a decrease in power in transmission from one point to another," will be encountered.

Maximum possible power transfer between two circuits has been shown to occur when the impedances involved were conjugates. The **transition loss** in transferring power from one circuit to another is a comparison of the power that is transferred under actual circuit conditions to the power that would be transferred if the load impedance and the source were conjugates.

Transition loss is not widely used in studying circuit performance. One reason is that communication equipment is built in a particular way for reasons other than impedance considerations, and the inherent

impedances of the equipment often cannot be altered at will. Also, as will be made clear in Chapters 6 and 7, lines and cables should be terminated in load impedances having the same magnitudes and angles as their characteristic impedances.

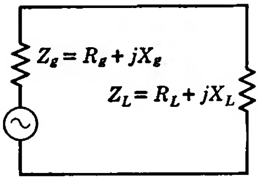


FIG. 20. Circuit for studying reflection loss.

Reflection Loss. Reflection loss and not transition loss is commonly used in measuring the loss in transferring power from one circuit or device to another. The reflection loss is zero when the input impedance Z_L of the circuit or device receiving the power equals in both magnitude and angle (including sign) the internal impedance Z_g of the circuit or device delivering the power.

In Fig. 20 is shown a generator supplying power to a load impedance. The power delivered to the load impedance is

$$P_L = I^2R_L = \frac{E_g{}^2R_L}{(Z_g + Z_L)^2}. \tag{50}$$

For no reflection loss, $R_L = R_g$, and $X_L = X_g$. For this condition the power P_L' delivered to the load impedance is

$$P_L' = I^2R_L = \frac{E_g{}^2R_L}{(2Z_L)^2} = \frac{E_g{}^2R_L}{(2Z_g)^2} = \frac{E_g{}^2R_g}{4Z_g{}^2}. \tag{51}$$

The *power reflection loss* is determined from the ratio of the power transferred to the load impedance under actual circuit conditions to the power that would be transferred under the prescribed conditions for no reflection loss. Dividing equation 50 by equation 51 gives

$$\frac{P_L}{P_L'} = \frac{4Z_g{}^2}{(Z_g + Z_L)^2} \times \frac{R_L}{R_g} = \frac{4Z_gZ_L}{(Z_g + Z_L)^2} \times \frac{\cos\theta_L}{\cos\theta_g}, \tag{52}$$

because $R_L = Z_L \cos\theta_L$, and $R_g = Z_g \cos\theta_g$. Equation 52 may be written

$$\frac{P_L}{P_L'} = k^2 \frac{\cos\theta_L}{\cos\theta_g}, \tag{53}$$

where $k = \dfrac{\sqrt{4Z_gZ_L}}{Z_g + Z_L}$, and sometimes is called the **reflection factor.**[4]

The **reflection loss** is defined[1] as

$$\text{Reflection loss} = 20 \log_{10} \frac{1}{k} = 20 \log_{10} \left| \frac{Z_g + Z_L}{\sqrt{4Z_gZ_L}} \right|, \tag{54}$$

where the reflection loss is in decibels. This equation also can be

derived on the basis of the currents flowing under matched and mismatched conditions.[4] The two vertical lines enclosing a portion of equation 54 indicate that in finding the reflection loss the impedances are combined as complex numbers and that the log is taken of the magnitude of the final value of impedance.

It is common practice to state that two connected circuits are **matched** when the input impedance of the driven circuit equals the internal impedance of the driving circuit in both magnitude and angle (including sign). Because the condition for matched impedances is not the condition for maximum power transfer (which is that the circuits are conjugate) it is possible to have **reflection gains.** Reflection losses and gains can be determined from Fig. 21.

Insertion Loss. In a communication transmission system, the magnitude of the loss caused by the *insertion* into the system of a new network or device must often be known. This loss is determined by several factors, such as the degree of impedance match at the input terminals, the internal transmission loss in the device to be inserted, and the degree of impedance match at the output terminals.

By definition,[1] the "**insertion loss** at a given frequency caused by the insertion of apparatus in a transmission system is the ratio, expressed in decibels, of the powers at that frequency delivered to that part of the system beyond the point of insertion before and after the insertion."

A generalized equation that will give the insertion loss of any line, circuit, or device becomes quite involved and will not be given.[4, 5] As an example, consider the insertion loss when a network, such as a few miles of telephone cable having a transmission loss (or attenuation loss, Chapter 7) within itself of L_N decibels, is inserted between a generator and a load. For these conditions, four losses will be involved in determining the insertion loss. These are the reflection loss L_1 at the junction of the network and the generator; the reflection loss L_2 at the junction of the network and the load impedance; the loss L_N within the cable (or network); and the reflection loss L_3 which would occur if the load impedance were connected directly to the generator. For this example the insertion loss is $L_1 + L_2 + L_N - L_3$, the loss L_3 being subtracted because it enters into determining the received power both before and after the insertion of the cable or other network.

Pads and Attenuators. It is often desired in communication circuits to reduce the amplitude of a signal wave. For example, the program signal delivered over a telephone line may be transmitted at a power level sufficient to "override" line noise (Chapter 14). A **pad,** defined[1] as "a non-adjustable transducer for reducing the amplitude

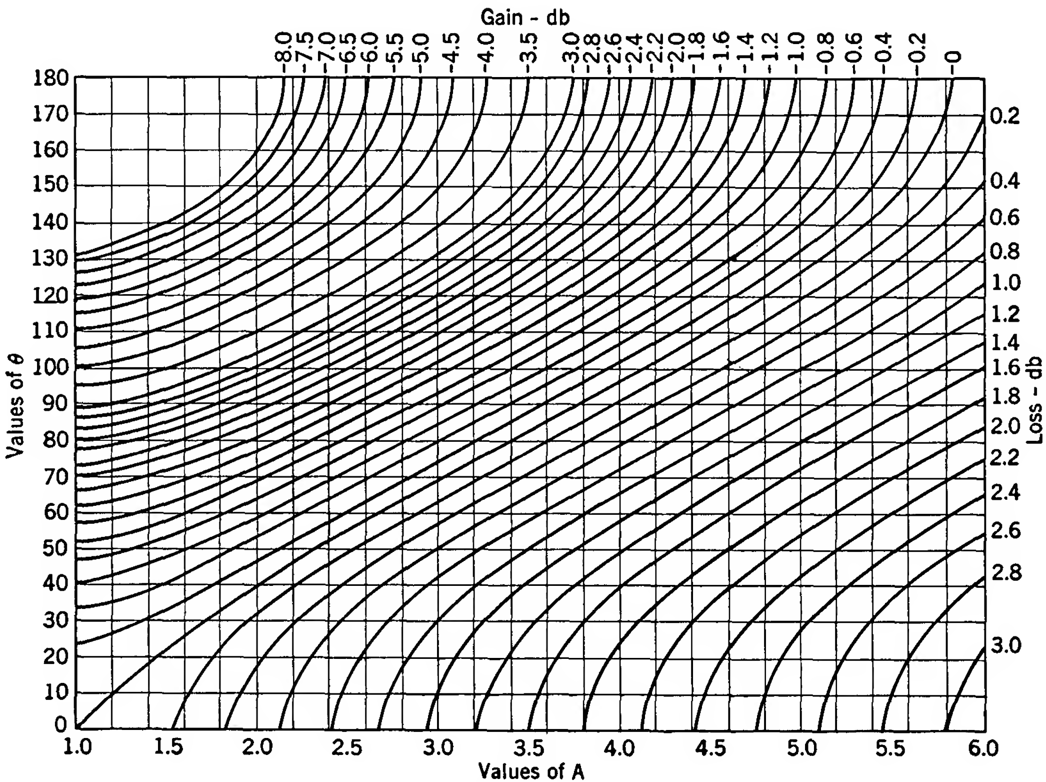


FIG. 21. Chart for determining reflection losses when two circuits having different impedances are connected.

Notes:

Values of the reflection loss for any two impedances Z_x and Z_y are here given as a function of the ratio

$$\frac{Z_x}{Z_y} = A\underline{/\theta}$$

where *A is the ratio of the magnitudes* and θ is the difference between the angles of the two impedances. The ratio is always to be taken by dividing the larger impedance by the smaller, i.e., so that A will not be less than unity. It is immaterial whether θ is positive or negative.

Negative values of reflection losses are reflection gains.

Examples:

(1) Find reflection loss for the impedances $200\underline{/60°}$ and $700\underline{/30°}$

$$A\underline{/\theta} = \frac{700\underline{/30°}}{200\underline{/60°}} = 3.5\,\underline{/-30°}$$

Reflection loss = 1.4 db.

(2) Find reflection loss for the impedances $400\overline{/35°}$ and $680\underline{/45°}$

$$A\underline{/\theta} = \frac{680\underline{/45°}}{400\overline{/35°}} = 1.7\underline{/80°}$$

Reflection loss = − 1.8 db.
(a gain)

of a wave without introducing appreciable distortion," may be necessary at the receiving end of the line to reduce the signal volume to the desired amount. An **attenuator,** defined[1] as an "adjustable transducer for reducing the amplitude of a wave without introducing appreciable distortion," is used in radio speech-input equipment to vary the volume of the program signal.

Pads are made of fixed resistors and will function over a wide frequency range, determined by the high-frequency characteristics of the resistors. Pads are commonly made in the form shown in Fig. 22, particularly if they are to be used with balanced circuits. In the design of pads, the facts known are usually the impedances of the devices with which the pad is to be associated and the loss in decibels that the pad is to have.

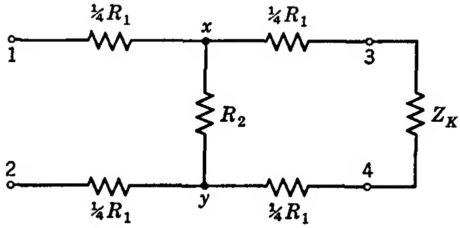


FIG. 22. Circuit for studying the design of a balanced pad. For a pad, the iterative impedance Z_K is resistance, often designated R_K.

In accordance with the theory on page 156, the iterative impedance of the pad of Fig. 22 is

$$R_K = \sqrt{R_1R_2 + \frac{R_1{}^2}{4}}, \tag{55}$$

the letter R_K being used because the iterative impedance will be resistance.

The power loss in decibels introduced by the pad can be found from power, voltage, or current ratios. If a voltage E_{12} is impressed at the sending end, the input current will be $I_{12} = E_{12}/R_K$, because the pad is symmetrical and is terminated in R_K. The voltage across the junctions x–y will be

$$E_{xy} = E_{12} - \tfrac{1}{2}R_1I_{12} = E_{12} - \frac{\frac{1}{2}E_{12}R_1}{R_K} = E_{12}\left(1 - \frac{\frac{1}{2}R_1}{R_K}\right). \tag{56}$$

The current through R_K will be $E_{xy}/(\frac{1}{2}R_1 + R_K)$, and this current multiplied by R_K will give the output voltage E_{34} across R_K. Thus,

$$E_{34} = \frac{E_{12}(1 - \frac{1}{2}R_1/R_K)R_K}{\frac{1}{2}R_1 + R_K} = \frac{E_{12}(R_K - \frac{1}{2}R_1)}{R_K + \frac{1}{2}R_1},$$

and (56*a*)

$$\frac{E_{12}}{E_{34}} = \frac{R_K + \frac{1}{2}R_1}{R_K - \frac{1}{2}R_1}.$$

As an illustration of the use of this equation, suppose that a pad such as Fig. 22 is to introduce a loss of 10 decibels and must have an iterative impedance of 600 ohms. The voltage ratio $E_{12}/E_{34} = 10^{0.05\times 10} = 3.162$, and, from equation 56*a*, the value of R_1 will be 622.5 ohms. With R_K and R_1 both known, from equation 55 $R_2 = 422$ ohms. Thus the pad of Fig. 22 should be composed of four 156-ohm resistors, and one 422-ohm resistor. Sometimes unsymmetrical **taper pads**[6] having different image impedances are used between circuits of different iterative impedances.

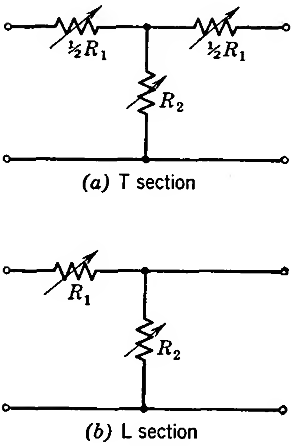


FIG. 23. Unbalanced attenuators commonly used.

Attenuators are made of variable resistors, usually arranged so that they are varied by a common knob. An attenuator of the configuration shown in Fig. 22 is a symmetrical, balanced device. The various resistances that an attenuator must have can be calculated as just explained for each different attenuator setting in decibels. Or calculations can be made at several settings, and curves can be plotted from which the values at other settings can be obtained.

An unbalanced pad commonly used is shown in Fig. 23*a*. This pad also can be designed as just explained. The L-section attenuator of Fig. 23(*b*) is widely used in unbalanced circuits where the input and the output impedances offered by the attenuator need not remain fixed. This is an unsymmetrical device, but it is satisfactory for many purposes. Equations for the design of this attenuator can be derived in the same way as for the symmetrical pad.

Propagation Constant of a Transducer. As an electromagnetic wave, such as an audio-frequency signal wave, travels along a network such as Fig. 6, two important phenomena occur. *First,* the magnitudes of the voltage and current at the input terminals will not be the same as the magnitudes of the voltage and current at the output terminals. For a network such as Fig. 6 the output voltages and currents will be less than the input voltages and currents, and **attenuation,** defined[1] as "a decrease in amplitude," will have occurred. *Second,* because time is required for a signal wave to travel through a network, the output voltage will not be in phase with the input voltage, and the output current will not be in phase with the input current. Thus, a **time delay** and a **phase shift** will occur.

The magnitudes of the attenuation and phase shift that occur are determined by the **propagation constant** of a transducer, defined[1] as "the natural logarithm of the ratio of the current entering the transducer to the current leaving the transducer, when the transducer is terminated in its iterative impedances." This definition applies to symmetrical networks; corresponding definitions for unsymmetrical networks will be found in reference 1. The propagation constant is usually represented by γ and is a complex number. By definition,[1] the real part of the propagation constant is the **attenuation constant** (usually represented by α), and the imaginary part is the **phase constant** (usually represented by β).

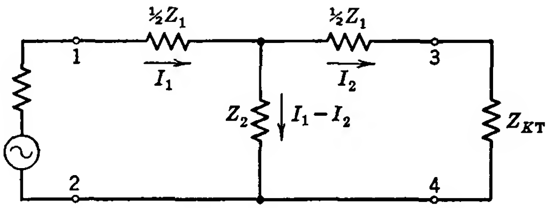


FIG. 24. Circuit for determining the propagation constant of a transducer.

If the current entering the network of Fig. 24 is I_1 and if the current leaving the network is I_2, the current through the shunt element will be $I_1 - I_2$. It can be written that

$$(I_1 - I_2)\, Z_2 = I_2\, (\tfrac{1}{2}Z_1 + Z_{KT})$$

and that

$$\frac{I_1}{I_2} = \frac{\dfrac{Z_1}{2} + Z_2 + Z_{KT}}{Z_2} \quad \text{and} \quad \frac{I_1}{I_2} = \frac{\dfrac{Z_1}{2} + Z_2 + \sqrt{Z_1 Z_2 + \dfrac{Z_1^{\,2}}{4}}}{Z_2}, \tag{57}$$

when the value of Z_{KT} from equation 45 is substituted. From the definition for propagation constant,

$$\gamma = \log_\epsilon \frac{I_1}{I_2} = \log_\epsilon \frac{\dfrac{Z_1}{2} + Z_2 + \sqrt{Z_1 Z_2 + \dfrac{Z_1^{\,2}}{4}}}{Z_2}$$

$$= \cosh^{-1}\left(1 + \frac{Z_1}{2Z_2}\right) = 2 \sinh^{-1} \frac{1}{2}\sqrt{\frac{Z_1}{Z_2}}. \tag{58}$$

The mathematical proof of these relations is given in detail in reference 5, on page 123, and in appendix C, from which the following has been taken:

Assume that $U = 1 + (Z_1/2Z_2)$. Then, equation 58 may be written

$$\gamma = \log_\epsilon (U + \sqrt{U^2 - 1}) = \cosh^{-1} U. \tag{59}$$

From fundamental mathematical relations,

$$\sinh x = \tfrac{1}{2} (\epsilon^x - \epsilon^{-x}), \text{ and } \cosh x = \tfrac{1}{2} (\epsilon^x + \epsilon^{-x}).$$

Then, from equation 59,

$$U = \cosh \gamma = \tfrac{1}{2} (\epsilon^\gamma + \epsilon^{-\gamma}). \tag{60}$$

$$U^2 - 1 = \tfrac{1}{4} (\epsilon^\gamma + \epsilon^{-\gamma})^2 - 1 = \tfrac{1}{4} (\epsilon^{2\gamma} - 2 + \epsilon^{-2\gamma}) = \tfrac{1}{4} (\epsilon^\gamma - \epsilon^{-\gamma})^2$$

$$U + \sqrt{U^2 - 1} = \tfrac{1}{2} (\epsilon^\gamma + \epsilon^{-\gamma}) + \tfrac{1}{2} (\epsilon^\gamma - \epsilon^{-\gamma}) = \epsilon^\gamma.$$

Hence,

$$\log_\epsilon (U + \sqrt{U^2 - 1}) = \log_\epsilon \epsilon^\gamma = \gamma = \cosh^{-1} U,$$

and when the value $U = 1 + (Z_1/2Z_2)$ is substituted, the first hyperbolic term of equation 58 is proved.

The proof that $\gamma = \cosh^{-1} U = 2 \sinh^{-1} \sqrt{\dfrac{U - 1}{2}}$ is as follows:

From equation 60,

$$U = \cosh \gamma = \tfrac{1}{2} (\epsilon^\gamma + \epsilon^{-\gamma}), \quad \text{or} \quad \frac{U - 1}{2} = \tfrac{1}{4} (\epsilon^\gamma - 2 + \epsilon^{-\gamma}).$$

Then,

$$\sqrt{\frac{U - 1}{2}} = \tfrac{1}{2} (\epsilon^{\gamma/2} - \epsilon^{-\gamma/2}) = \sinh \frac{\gamma}{2} \quad \text{and} \quad \gamma = 2 \sinh^{-1} \sqrt{\frac{U - 1}{2}}.$$

When the value of U is substituted, the last part of equation 58 is proved.

Electric Wave Filters. A filter is defined[1] as "a selective network which transmits freely electric waves having frequencies within one or more frequency bands and which attenuates substantially electric waves having other frequencies." Electric wave filters were invented by Campbell, a contribution of great importance.

Many types of filters are used in practice,[4, 5, 7] but it is feasible to consider only the simplest structures. These will be composed of inductors and capacitors only, so that they may transmit freely over certain bands of frequencies. The assumption is made that the inductors and capacitors are lossless, and that *no energy can be dissipated within the filter.* The treatment will be confined largely to ladder-type structures, such as the equivalent T and π sections of Fig. 6.

Of great importance in filter design is the width of the transmitted band of frequencies. There are several ways of determining this, two of which will now be given.

Transmitted Bands from Propagation Constant. As shown by equation 58, the propagation constant for a T section such as Fig. 24 is $\gamma = 2 \sinh^{-1} \frac{1}{2}\sqrt{Z_1/Z_2}$. Since the elements of a filter are assumed to be without resistance and to cause no energy losses, Z_1 and Z_2 must be pure reactances. Accordingly, the ratio Z_1/Z_2 must have an angle that is either 0 or 180°; that is, the ratio must be a *real* positive quantity, or a *real* negative quantity, but cannot be complex. For instance, if Z_1 and Z_2 are both inductors with negligible resistance, the ratio Z_1/Z_2 will have a zero angle; if Z_1 is inductive and Z_2 capacitive, then Z_1/Z_2 will have an angle of 180°. The effects of these different ratios of Z_1/Z_2 on the propagation constant will be considered now.[5]

Case I. When Z_1/Z_2 is positive. The propagation constant γ is composed of a real and an imaginary term; that is, $\gamma = \alpha + j\beta$. When Z_1/Z_2 of equation 58 gives a zero angle, γ is a *real number* since the $j\beta$ part becomes zero. There will accordingly be undesired attenuation in the filter for all positive values of Z_1/Z_2.

Case II. When Z_1/Z_2 is negative and less than -4. If the ratio Z_1/Z_2 has an angle of 180° it is negative in sign; and, therefore, $\frac{1}{2}\sqrt{Z_1/Z_2}$ of equation 58 has an angle of 90° and both $\frac{1}{2}\sqrt{Z_1/Z_2}$ and γ are pure imaginaries. Equation 58 can be written

$$\sinh\frac{\gamma}{2} = \sinh\left(\frac{\alpha}{2} + \frac{j\beta}{2}\right) = \frac{1}{2}\sqrt{\frac{Z_1}{Z_2}}. \tag{61}$$

From hyperbolic trigonometry, $\sinh (x + jy) = \sinh x \cos y + j \cosh x \sin y$. Thus,

$$\sinh\left(\frac{\alpha}{2} + j\frac{\beta}{2}\right) = \sinh\frac{\alpha}{2}\cos\frac{\beta}{2} + j\cosh\frac{\alpha}{2}\sin\frac{\beta}{2} = \frac{1}{2}\sqrt{\frac{Z_1}{Z_2}}. \tag{62}$$

Since it was previously shown that $\frac{1}{2}\sqrt{Z_1/Z_2}$ and γ are pure imaginaries for the case being considered, then the real part of equation 62, that is, $\sinh \alpha/2 \cos \beta/2$, must be zero. For this to be true, either $\sinh \alpha/2$ or $\cos \beta/2$ must equal zero. If $\sinh \alpha/2 = 0$, then $\cosh \alpha/2 = 1$, because of the numerical relations between these two hyperbolic functions. When these relations exist, equation 62 becomes

$$j\sin\beta/2 = \frac{1}{2}\sqrt{\frac{Z_1}{Z_2}},\ \sin\frac{\beta}{2} = \frac{1}{2}\sqrt{-\frac{Z_1}{Z_2}},\ \text{and } \beta = 2\sin^{-1}\frac{1}{2}\sqrt{-\frac{Z_1}{Z_2}}; \tag{63}$$

therefore,

$$\gamma = 0 + j2\sin^{-1}\frac{1}{2}\sqrt{-\frac{Z_1}{Z_2}}. \tag{64}$$

If Z_1/Z_2 is greater than -4, then $\frac{1}{2}\sqrt{-Z_1/Z_2}$ would exceed 1, and equation 63 cannot hold because the sine of an angle cannot exceed unity. Thus, from equation 64 a filter will transmit *without attenuation* when Z_1/Z_2 is negative and less than -4 in magnitude.

Case III. When Z_1/Z_2 is more negative than -4. For values of Z_1/Z_2 more negative than -4, the real part $\sinh \alpha/2 \cos \beta/2$ of equation 62 can be made zero by $\cos \beta/2 = 0$. At the values $\cos \beta/2 = 0$, $\sin \beta/2 = \pm 1$. Thus the angle $\beta/2$ must be some odd multiple of 90° to give these zero and unity values. Expressed in terms of radians (90° $= \pi/2$ radians),

$$\frac{\beta}{2} = (2K - 1)\frac{\pi}{2}, \quad \text{and} \quad \beta = (2K - 1)\pi, \tag{65}$$

where K is any whole number. This relation must hold to give $\cos \beta/2 = 0$ values.

When $\cos \beta/2 = 0$ and $\sin \beta/2 = \pm 1$ are placed in equation 62, this becomes

$$\cosh \frac{\alpha}{2} = \frac{1}{2}\sqrt{-\frac{Z_1}{Z_2}}. \tag{66}$$

From this relation

$$\frac{\alpha}{2} = \cosh^{-1} \frac{1}{2}\sqrt{-\frac{Z_1}{Z_2}}, \quad \text{and} \quad \alpha = 2 \cosh^{-1} \frac{1}{2}\sqrt{-\frac{Z_1}{Z_2}}. \tag{67}$$

Then, from equations 65 and 67, the propagation constant becomes

$$\gamma = 2 \cosh^{-1} \frac{1}{2}\sqrt{-\frac{Z_1}{Z_2}} + j\,(2K - 1)\,\pi. \tag{68}$$

This relation shows that, for negative values of Z_1/Z_2 greater than -4, the filter will offer attenuation.

From the foregoing, it appears that a non-dissipative recurrent structure of the type shown in Fig. 4 having series impedances Z_1 and shunt impedances Z_2 will pass readily only those signals of frequencies such that the ratio Z_1/Z_2 will lie between zero and -4. The values of Z_1 and Z_2 depend on the frequency because in filters Z_1 and Z_2 are inductors and capacitors.

Transmitted Band from Iterative Impedances. The frequencies transmitted by a filter can be determined from the iterative impedances of the filter sections by the following method. Assume that a resistance load is connected to a source through a filter composed of inductors and capacitors having negligible losses. Then, the filter *itself* cannot absorb power.

Now suppose that a frequency is chosen such that the iterative impedance of the filter is purely resistive. At this frequency, the source will send power to the filter, and, since the filter cannot absorb power, it must reach the load.

Next assume that at a different frequency the iterative impedance of the filter is purely reactive. This prevents the source from sending any power into the connected circuit, and thus no power reaches the load at this frequency. From these facts it may be concluded that a filter will transmit without attenuation those frequencies for which the iterative impedance Z_K is resistive (Fig. 25).

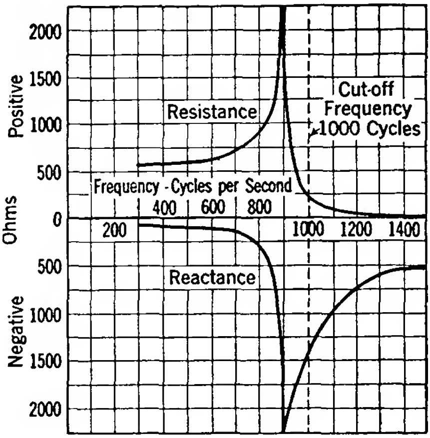


FIG. 25. Approximate resistance and reactance components of the iterative impedance of a typical low-pass filter.

The iterative impedance given by equation 45 is

$$Z_{KT} = \sqrt{Z_1 Z_2 + Z_1{}^2/4} = \sqrt{Z_1(Z_2 + Z_1/4)}.$$

For unattenuated transmission this must be a positive real number to be resistive. This equation consists of two parts: $X_a = Z_1$ and $X_b = (Z_2 + Z_1/4)$. Thus, $Z_{KT} = \sqrt{X_a X_b}$. Pure reactances only are being considered; and for Z_{KT} to be positive, if X_a represents inductive reactance and is $+jX_a$, then X_b must be capacitive reactance or $-jX_b$, and vice versa, *within the band of frequencies transmitted.* For these relations $Z_{KT} = \sqrt{(+jX_a)\ (-jX_b)}$, and Z_{KT} will be positive. To satisfy these conditions, Z_2 must be, *first,* the opposite type of reactance to Z_1, and *second,* greater in numerical value than $Z_1/4$. Or, in other words, the ratio $Z_1/4Z_2$ from equation 45 must lie between 0 and -1 for the frequencies to be transmitted, as shown in Fig. 26.

"Constant-k" Filters. Constant-k filters are sometimes used by themselves, and sometimes as a basis for the design of the other types. In the constant-k filter, Z_1 and Z_2 are inverse elements; that is, if Z_1 of Fig. 24 consists of inductance, then Z_2 must consist of capacitance. That is,

$$Z_1 Z_2 = k^2, \tag{69}$$

where k is a constant independent of frequency. Because of the fact

that the other filters are derived from it, the constant-k filter is known as a "prototype."

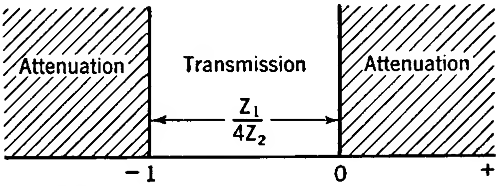


FIG. 26. Relations of Z_1 and Z_2 for unattenuated transmission.

Low-Pass Filter. A properly terminated low-pass filter with no power losses in the elements will pass, without attenuation, signals of all frequencies from zero up to a critical **cutoff frequency,** and will attenuate signals above this frequency. Three sections of a low-pass filter are shown in Fig. 27. These will pass signals of such frequencies that Z_1/Z_2 lies between 0 and −4.

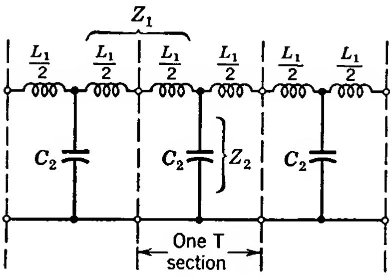


FIG. 27. Three constant-k low-pass filter sections.

In Fig. 27 the series impedances are $Z_1 = j\omega L_1$, and the shunt impedances are $Z_2 = 1/(j\omega C_2)$. Then, $Z_1/Z_2 = (j\omega L_1)(j\omega C_2) = -\omega^2 L_1 C_2$. When $Z_1/Z_2 = 0$, the lower cutoff frequency is found; thus, since $\omega = 2\pi f$, $Z_1/Z_2 = -\omega_c{}^2 L_1 C_2 = -(2\pi f_c')^2 L_1 C_2 = 0$, and $f_c' = 0$. When $Z_1/Z_2 = -\omega_c{}^2 L_1 C_2 = -(2\pi f_c'')^2 L_1 C_2 = -4$, then the upper cutoff frequency is

$$f_c'' = \frac{1}{\pi\sqrt{L_1 C_2}}. \tag{70}$$

The low-pass filter will accordingly pass currents of all frequencies from direct current ($f = 0$) up to the value given by equation 70.

The variations of the iterative impedances of T and π low-pass filter sections will now be investigated. From equation 45,

$$Z_{KT} = \sqrt{Z_1 Z_2\left[1 + \frac{1}{4}\frac{Z_1}{Z_2}\right]} = \sqrt{\frac{j\omega L_1}{j\omega C_2}\left[1 + \frac{1}{4}\frac{j\omega L_1}{\frac{1}{j\omega C_2}}\right]}$$

$$= \sqrt{\frac{L_1}{C_2}}\sqrt{1 - \frac{L_1 C_2 \omega^2}{4}} = \sqrt{\frac{L_1}{C_2}}\sqrt{1 - \left(\frac{f}{f_c''}\right)^2}, \tag{71}$$

when $-\omega_c^2 L_1 C_2 = -4$ and $\omega_c'' = 2\pi f_c''$ are substituted. This equation gives the iterative impedance of the low-pass T section of Fig. 27 at any given frequency ratio f/f_c''. By a similar process it can be shown that, for the π section, equation 46 becomes

$$Z_{K\pi} = \frac{\sqrt{Z_1 Z_2}}{\sqrt{1 + \frac{1}{4}\frac{Z_1}{Z_2}}} = \frac{\sqrt{\frac{j\omega L_1}{j\omega C_2}}}{\sqrt{1 + \frac{1}{4}\frac{j\omega L_1}{\frac{1}{j\omega C_2}}}}$$

$$= \frac{\sqrt{L_1/C_2}}{\sqrt{1 - \frac{1}{4}\omega^2 L_1 C_2}} = \frac{\sqrt{L_1/C_2}}{\sqrt{1 - \left(\frac{f}{f_c''}\right)^2}}. \quad (72)$$

The equations for these two impedance curves are shown plotted in Fig. 28. They show that, for $f = 0$,

$$Z_K = \sqrt{L_1/C_2}, \quad (73)$$

for both the T and the π sections. They further show that the impedance values then diverge and that at the value $f = f_c''$ (that is, at the cutoff frequency) $Z_{KT} = 0$ and $Z_{K\pi} = \infty$. It is common practice to design a low-pass filter so that the iterative impedance Z_K computed at $f = 0$ and given by equation 73 equals the impedance of the line in which it is to operate.

It is now possible to derive the *design equations* for low-pass filters. The upper cutoff frequency was found from the relation $-\omega_c^2 L_1 C_2 = -4$. If $2\pi f_c''$ is substituted for ω_c, this relation can be arranged $L_1/C_2 = (2\pi f_c'' L_1)^2/4$. But $Z_K = \sqrt{L_1/C_2}$, and thus

$$Z_K = \pi f_c'' L_1 \quad \text{and} \quad L_1 = Z_K/(\pi f_c'') \text{ henrys.} \quad (74)$$

Therefore,

$$Z_K = \sqrt{\frac{Z_K/(\pi f_c'')}{C_2}}, \quad Z_K = \frac{1}{\pi f_c'' C_2}, \quad \text{and} \quad C_2 = \frac{1}{\pi f_c'' Z_K} \text{ farads.} \quad (75)$$

Using these two equations, simple low-pass constant-k filters can be designed as will be illustrated.

Suppose that it is desired to construct a low-pass filter having an iterative impedance of 600 ohms and a cutoff frequency of 1000 cycles.

From equation 74, $L_1 = 600/(3.1416 \times 1000) = 0.191$ henry, and from equation 75, $C_2 = 1/(3.1416 \times 1000 \times 600) = 0.000000530$ farad or 0.53 microfarad. That is, the series element Z_1 of Fig. 27 must be 0.191 henry inductance and the shunt element Z_2 of Fig. 27 must be 0.53 microfarad capacitance.

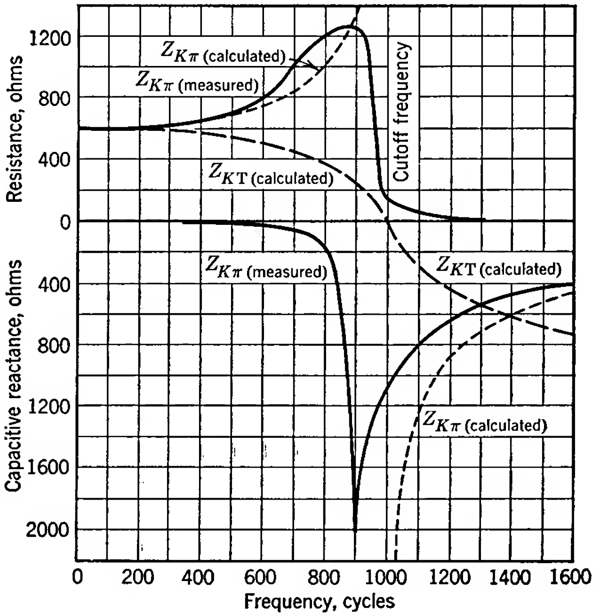


FIG. 28. Theoretical curves are shown by broken lines for lossless low-pass constant-k filters. These are plots of equations 71 and 72 for 600-ohm filters having the values calculated on this page. The solid curves are the resistance and reactance components as measured on an actual π section having the values calculated on this page. The measurements were made with an impedance bridge, and the impedance found from the relation $Z_K = \sqrt{Z_{oc}Z_{sc}}$.

The filter section can be connected either as a T or as a π section. If connected as a T section, two inductors each having an inductance of 0.191/2 or 0.0955 henry, and one capacitor having a capacitance of 0.53 microfarad, will be required. If connected as a π section, one 0.191-henry inductor and two $0.53/2 = 0.265$-microfarad capacitors will be needed. These relations will be evident from Figs. 4, 5, and 6. It should also be noted that

$$Z_K = \sqrt{L_1/C_2} = \sqrt{\frac{0.191}{0.00000053}} = 600 \text{ ohms (approximately).}$$

Of particular interest in filter studies are the iterative impedance that has been considered, and the attenuation and phase shift that will now be treated. If the filter elements are pure reactances, the attenuation will be zero for frequencies within the band passed and will be as given by equation 67 for frequencies beyond the band passed. For the low-pass filter of Fig. 27,

$$\alpha = 2\cosh^{-1}\frac{1}{2}\sqrt{-\frac{Z_1}{Z_2}} = 2\cosh^{-1}\frac{1}{2}\sqrt{-\left[\frac{j2\pi f L_1}{-j\dfrac{1}{2\pi f C_2}}\right]} =$$

$$2\cosh^{-1}\tfrac{1}{2}\sqrt{(2\pi f)^2 L_1 C_2} = 2\cosh^{-1}\frac{f}{f_c''}, \tag{76}$$

when the value given by equation 74 is substituted for L_1, and the value given by equation 75 is substituted for C_2. Thus, at 2000 cycles the attenuation of the low-pass filter under consideration will be $\alpha = 2\cosh^{-1}(2000/1000) = 2\cosh^{-1} 2$, and $\alpha = 2 \times 1.31 = 2.62$ nepers, the value 1.31 being determined from a table of hyperbolic functions. The corresponding loss in decibels is db $= 2.62 \times 8.686 = 22.8$ decibels. The dotted curve of Fig. 29 was computed in this way. The solid curve of Fig. 29 was measured on an actual filter section, the elements of which had some loss. For methods of calculating the attenuation in which the losses of the filter elements are considered, reference 4 should be consulted.

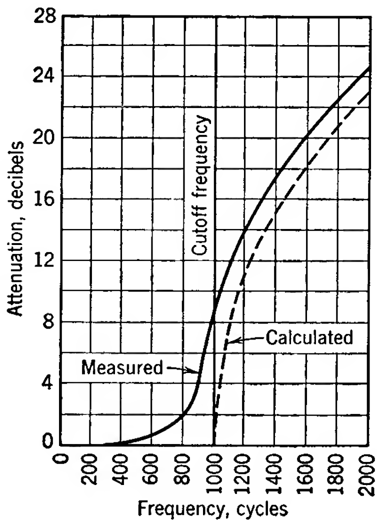


FIG. 29. Attenuation curves as calculated using equation 76 and as measured. These curves are for one π section of the constant-k low-pass filter calculated on page 170. Small variations in the measured values below cutoff have been neglected.

If the elements of a low-pass filter are pure reactances, there will be a varying phase shift over the band passed, and a constant phase shift of 180° beyond the band passed. From equation 63, and for the low-pass filter,

$$\beta = 2\sin^{-1}\frac{1}{2}\sqrt{-\frac{Z_1}{Z_2}} = 2\sin^{-1}\frac{f}{f_c''}, \tag{77}$$

when the substitutions as in equation 76 are made. Thus, for the low-pass filter under consideration, and at 500 cycles, $\beta = 2 \sin^{-1} (500/1000) = 2 \sin^{-1} 0.5 = 60°$. The dotted curve of Fig. 30 was

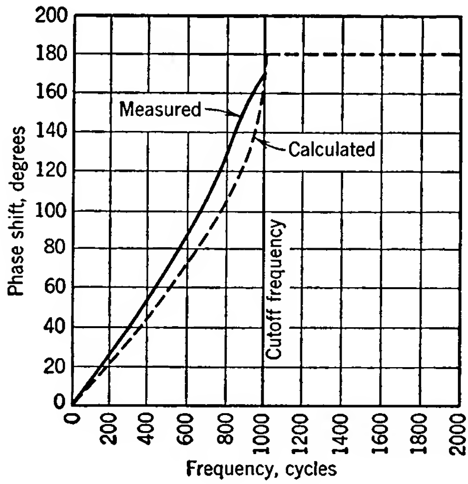


FIG. 30. Phase-shift curves for one π section of a constant-k low-pass filter as calculated by equation 77, and as measured with a cathode-ray oscilloscope. The input voltage was impressed on one set of deflecting plates, the output was impressed on the other set, and the phase angle was determined from the resulting figure. (Page 310.)

computed in this manner, and the solid curve was measured on an actual filter section, the elements of which had some loss. The phase measurements were made with a cathode-ray oscilloscope (page 309).

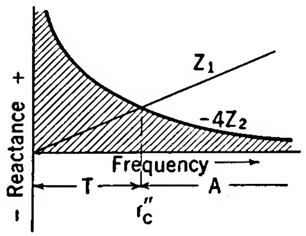


FIG. 31. Reactance curves for a low-pass filter.

The transmission and attenuation bands of low-pass filters can be determined from simple reactance sketches as in Fig. 31. Thus, if curves for Z_1 (equals $2\pi L_1$) and $-4Z_2$ (where $Z_2 = 1/(2\pi f C_2)$) are plotted as indicated, the transmission band T will occur where Z_1 lies between the frequency axis and the curve $-4Z_2$, and the band A where Z_1 lies without curve $-4Z_2$ will be attenuated.

High-Pass Filters. Such a network freely passes all signals above a certain critical or cutoff frequency and greatly attenuates signals

below this frequency. Three T sections of a constant-k high-pass filter are shown in Fig. 32.

In the high-pass filter $Z_1 = 1/(j\omega C_1)$, since the condensers $2C_1$ are in series, $Z_2 = j\omega L_2$, and the ratio of these two impedances is $Z_1/Z_2 = -1/(\omega^2 L_2 C_1)$. From the general filter theory, the frequency

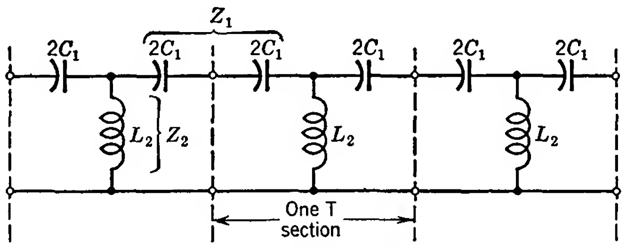


FIG. 32. Three high-pass filter sections.

band passed will lie between $Z_1/Z_2 = 0$ and $Z_1/Z_2 = -4$. When the first value is substituted, $Z_1/Z_2 = -1/(\omega^2 C_1 L_2) = 0$, and f_c'' will be infinite. When the second substitution is made, then $-1/(\omega^2 L_2 C_1) = -4$, and

$$f_c' = \frac{1}{4\pi\sqrt{L_2 C_1}}. \tag{78}$$

Thus, the high-pass filter of Fig. 32 will freely pass all frequencies between f_c' and infinity and will greatly attenuate all frequencies below f_c'.

The iterative impedance equations for the T section of Fig. 32 are found in the same general manner as for the low-pass filter. From equation 45,

$$Z_{KT} = \sqrt{Z_1 Z_2\left[1 + \frac{1}{4}\frac{Z_1}{Z_2}\right]} = \sqrt{\left(\frac{1}{j\omega C_1}\right)\left(j\omega L_2\right)\left[1 + \frac{1}{4}\frac{\frac{1}{j\omega C_1}}{j\omega L_2}\right]}$$

$$= \sqrt{\frac{L_2}{C_1}}\sqrt{1 - \frac{1}{4\omega^2 L_2 C_1}} = \sqrt{\frac{L_2}{C_1}}\sqrt{1 - \left(\frac{f_c'}{f}\right)^2}. \tag{79}$$

In a similar manner the iterative impedance for a π section is found from equation 46 to be

$$Z_{K\pi} = \frac{\sqrt{L_2/C_1}}{\sqrt{1 - \left(\frac{f_c'}{f}\right)^2}}. \tag{80}$$

As equations 79 and 80 indicate, Z_K varies with frequency in both the T and the π sections (see also Fig. 33). For both types of sections, $Z_K = \sqrt{L_2/C_1}$ at $f = \infty$. For the T section, $Z_{KT} = 0$ at the cutoff frequency; and for the π section, $Z_{K\pi} = \infty$ at the cutoff frequency.

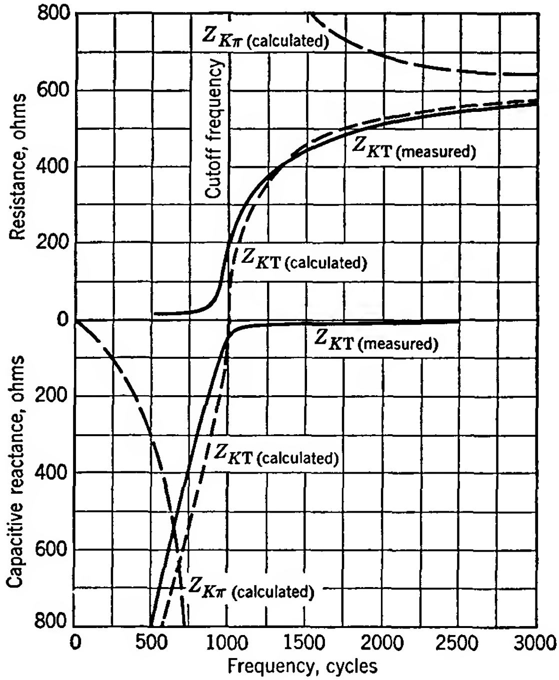


FIG. 33. Theoretical curves are shown by broken lines for lossless high-pass constant-k filters. These are plots of equations 79 and 80 for 600-ohm filters having the values calculated on page 175. The solid curves are the resistance and reactance components as measured on an actual T section having the values calculated on page 175. (See explanation accompanying Fig. 28 for method of making measurements.)

The iterative impedance used for high-pass filter design is at $f = \infty$; hence, for both the T and the π sections,

$$Z_K = \sqrt{L_2/C_1}. \tag{81}$$

The design formulas are now easily found. Since $Z_K = \sqrt{L_2/C_1}$, $1/(\omega^2 L_2 C_1) = 4$, and $L_2/C_1 = 4\omega_c^2 L_2^2$, then $Z_K^2 = 4\omega_c^2 L_2^2$, and therefore

$$L_2 = Z_K/(2\omega_c) = Z_K/(4\pi f_c') \text{ henrys}. \tag{82}$$

The value of C_1 can now be readily found as follows

$$Z_K = \sqrt{\frac{L_2}{C_1}}, \text{ and } Z_K{}^2 = \frac{L_2}{C_1} = \frac{\dfrac{Z_K}{4\pi f_c'}}{C_1} = \frac{Z_K}{4\pi f_c' C_1}, \text{ and } C_1 = \frac{1}{4\pi f_c' Z_K}. \quad (83)$$

To consider a typical case, suppose that it is desired to design a high-pass filter to operate between circuits of 600 ohms impedance, and to cut off at 1000 cycles. Then, $L_2 = 600/(4 \times 3.1416 \times 1000) = 0.0477$ henry, and $C_1 = 1/(4 \times 3.1416 \times 1000 \times 600) = 0.000000132$ farad or 0.132 microfarad. If Z_K were desired, and L_2 and C_1 had been given, then $Z_K = \sqrt{L_2/C_1} = \sqrt{0.0477/0.000000132} = 600$ ohms, approximately. As shown in Fig. 32, two capacitors of $2C_1$ and an inductor L_2 would be used for a T section. For a π section, two inductors of $2L_2$ and one capacitor of C_1 would be used.

The attenuation of a constant-k high-pass filter can be calculated from equation 67. Thus,

$$\alpha = 2 \cosh^{-1} \frac{1}{2}\sqrt{-\frac{Z_1}{Z_2}} = 2 \cosh^{-1} \frac{1}{2}\sqrt{-\left[\frac{-j\dfrac{1}{2\pi f C_1}}{j2\pi f L_2}\right]}$$

$$= 2 \cosh^{-1} \frac{1}{2}\sqrt{\frac{1}{(2\pi f)^2 L_2 C_1}} = 2 \cosh^{-1} \frac{f_c'}{f}, \quad (84)$$

when the value given by equation 82 is substituted for L_2, and the value given by equation 83 is substituted for C_1. At 500 cycles, for the high-pass filter under consideration, $\alpha = 2 \cosh^{-1} (1000/500) = 2 \cosh^{-1} 2$, and $\alpha = 2 \times 1.31 = 2.62$ nepers, or $2.62 \times 8.686 = 22.8$ decibels. The dotted curve of Fig. 34 shows calculated values of attenuation for the filter under consideration (assumed to be composed of lossless elements), and the solid curve shows the attenuation measured on an actual filter, designed as explained in the preceding paragraph.

The phase shift of a constant-k high-pass filter can be found from equation 63

$$\beta = 2 \sin^{-1} \frac{1}{2}\sqrt{-\frac{Z_1}{Z_2}} = 2 \sin^{-1} \frac{f_c'}{f}, \quad (85)$$

when the same substitutions as in equation 84 are made. For the filter under consideration, and at a frequency of 2000 cycles, $\beta = 2 \sin^{-1}(1000/2000) = 2 \sin^{-1} 0.5 = 60°$. The dotted curve of Fig. 35 shows calculated values, and the solid curve shows values measured with a cathode-ray oscilloscope.

As for the low-pass filter, an analysis of the characteristics of a high-pass filter can be made from simple reactance sketches as in Fig. 36. In this figure $Z_1 = 1/(2\pi f C_1)$ and $Z_2 = 2\pi f L_2$, giving the curves positions different from Fig. 31. Transmission will occur for such frequencies that Z_1 lies between $-4Z_2$ and the frequency axis as indicated.

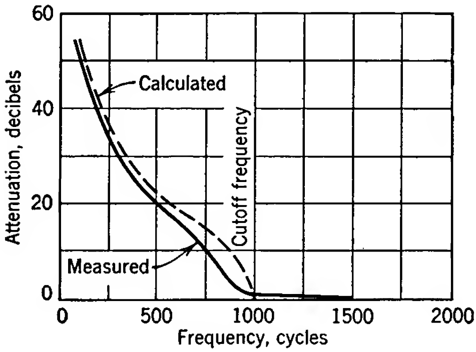


FIG. 34. Attenuation curves as calculated, using equation 84, and as measured. These curves are for one T section of the constant-k high-pass filter calculated on page 175. Small variations in the measured values above cutoff have been neglected.

Band-Pass Filter. This filter is designed to pass a given band of frequencies, and to attenuate all other frequencies. A band-pass filter of the constant-k type is shown in Fig. 37. The arm Z_1 is resonant for some frequency f_c, and Z_2 is anti-resonant (page 65) at this same value. At the resonant frequency f_r, Z_1 offers zero impedance and Z_2 offers an infinite impedance. Although more generalized structures are possible,[4, 5] only the simple "confluent" band-pass filter passing only one band will be considered. In this structure, $L_1C_1 = L_2C_2$.

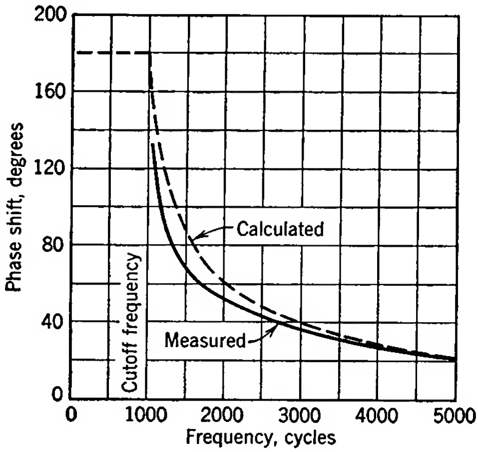


FIG. 35. Phase-shift curves for one T section of a constant-k high-pass filter as calculated by equation 85 and as measured as explained for Fig. 30. Constants calculated on page 175.

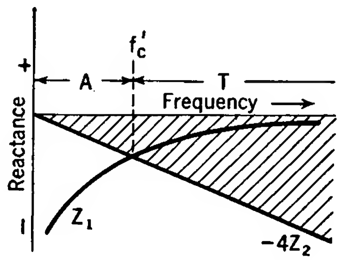


FIG. 36. Reactance curves for a high-pass filter.

For the filter of Fig. 37, $Z_1 = j\left(\omega L_1 - \dfrac{1}{\omega C_1}\right) = j\,\dfrac{(\omega^2 L_1 C_1 - 1)}{\omega C_1}$,

and $Z_2 = j\left(\dfrac{1}{\dfrac{1}{\omega L_2} - \omega C_2}\right) = j\left(\dfrac{\omega L_2}{1 - \omega^2 L_2 C_2}\right)\cdot$ The ratio Z_1/Z_2 (when $L_1C_1 = L_2C_2$, and $L_2 = L_1C_1/C_2$ are substituted) is

$$\frac{Z_1}{Z_2} = \frac{(\omega^2 L_1 C_1 - 1)^2}{-\omega^2 L_2 C_1} = \frac{(\omega^2 L_1 C_1 - 1)^2}{-\omega^2 \dfrac{L_1 C_1{}^2}{C_2}}. \tag{86}$$

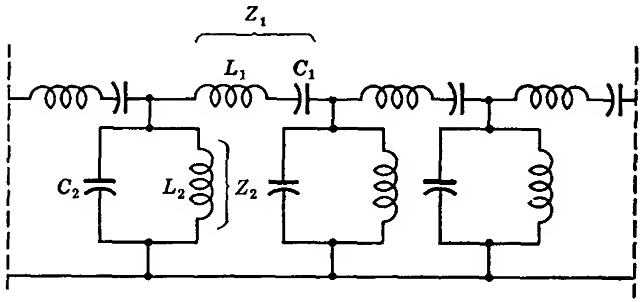


FIG. 37. A band-pass filter of the constant-k type.

A network of this general type will pass frequencies between $Z_1/Z_2 = -4$ and $Z_1/Z_2 = 0$. Accordingly, equation 86 becomes

$$\frac{(\omega_c{}^2 L_1 C_1 - 1)^2}{\dfrac{\omega_c{}^2 L_1 C_1{}^2}{C_2}} = 4, \quad \text{and} \quad \omega_c = \sqrt{\frac{1}{L_1C_1} + \frac{1}{L_1C_2}} \pm \frac{1}{\sqrt{L_1C_2}}. \tag{87}$$

As equation 87 indicates, an upper and a lower cutoff frequency are obtained; that is,

$$\left.\begin{aligned} f_c'' &= \frac{1}{2\pi}\left(\sqrt{\frac{1}{L_1C_1} + \frac{1}{L_1C_2}} + \frac{1}{\sqrt{L_1C_2}}\right) \\ \text{and} \qquad & \\ f_c' &= \frac{1}{2\pi}\left(\sqrt{\frac{1}{L_1C_1} + \frac{1}{L_1C_2}} - \frac{1}{\sqrt{L_1C_2}}\right). \end{aligned}\right\} \tag{88}$$

These cutoff points were found by $Z_1/Z_2 = -4$. When the ratio

is equated to zero,

$$\left.\begin{aligned} \frac{Z_1}{Z_2} &= \frac{(\omega_c^2 L_1 C_1 - 1)^2}{-\omega_c^2 L_2 C_1} = 0. \\ \text{Then,}\qquad \omega_c &= \frac{1}{\sqrt{L_1 C_1}} \quad \text{and} \quad f_c = \frac{1}{2\pi\sqrt{L_1 C_1}} = \frac{1}{2\pi\sqrt{L_2 C_2}}, \end{aligned}\right\} \tag{89}$$

since $L_1C_1 = L_2C_2$. This is not a true cutoff frequency,[4] but lies at the middle of the band passed. It is the frequency for which the series arm is resonant and the shunt arm antiresonant as previously considered. In a *generalized* filter there are two bands passed; but with $L_1C_1 = L_2C_2$ these two bands "flow together" and the filter is of the *simple confluent type.* If the two expressions of equation 88 are subtracted,

$$(f_c'' - f_c') = \frac{1}{\pi\sqrt{L_1 C_2}} = \frac{1}{\pi\sqrt{L_1^2 \dfrac{C_1}{L_2}}},$$

since $L_1C_1 = L_2C_2$. Then,

$$L_1 = \frac{1}{\pi\sqrt{\dfrac{C_1}{L_2}}\,(f_c'' - f_c')} = \frac{Z_K}{\pi(f_c'' - f_c')}, \tag{90}$$

where $Z_K = \sqrt{L_2/C_1} = \sqrt{L_1/C_2}$ and is the iterative impedance taken at the resonant frequency. By making similar substitutions,

$$C_1 = \frac{(f_c'' - f_c')}{4\pi f_c' f_c'' Z_K}, \quad L_2 = \frac{(f_c'' - f_c')Z_K}{4\pi f_c' f_c''}, \text{ and } C_2 = \frac{1}{\pi(f_c'' - f_c')Z_K}. \tag{91}$$

As an illustration of the design of a simple confluent constant-k band-pass filter, assume that it is desired to construct a filter such that $Z_K = 600$ ohms, $f_c' = 900$ cycles, and $f_c'' = 1100$ cycles. Then,

$$L_1 = \frac{600}{3.1416\ (1100 - 900)} = 0.955 \text{ henry},$$

$$L_2 = \frac{(1100 - 900)\ 600}{4 \times 3.1416 \times 900 \times 1100} = 0.00965 \text{ henry};$$

$$C_1 = \frac{(1000 - 900)}{4 \times 3.1416 \times 900 \times 1100 \times 600} = 0.0000000268 \text{ farad},$$

$$C_2 = \frac{1}{3.1416\ (1100 - 900)\ 600} = 0.00000265 \text{ farad}.$$

The values here computed are for the entire series and parallel arms of Fig. 37. If a T section is desired, it should be constructed as indicated in Fig. 38. Also, the values for a π section may be readily determined.

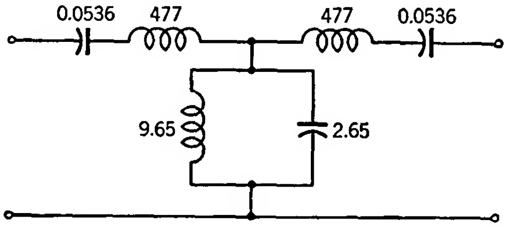


FIG. 38. A T section of a band-pass filter. Capacitance in microfarads. Inductance in millihenrys.

The attenuation curve for a band-pass filter of this type would be approximately as shown in Fig. 39.

The pass bands can also be computed by reactance sketches[4, 5] as in Fig. 40. This sketch is *not* for the simple confluent type just considered. It will therefore be noted that $-4Z_2$ values determined by the parallel or antiresonant circuit do not pass through infinity at the same point that the Z_1 curve determined by the series elements passes through zero. The two transmission bands T and T' are accordingly not confluent and, as the diagram indicates, exist separately. This figure represents a double band-pass filter.

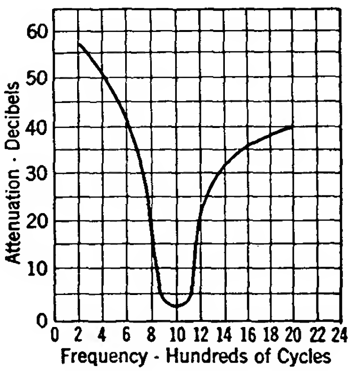


FIG. 39. Attenuation curve for three constant-k band-pass filter sections such as Fig. 38. Terminated in 600 ohms resistance.

The Band-Elimination Filter. A filter of this type is designed to pass freely currents of all frequencies except those within a definite band. Three sections of a constant-k band-elimination filter are shown in Fig. 41. By comparison with the band-pass filter,

$$Z_1 = j\left(\frac{\omega L_1}{1 - \omega^2 L_1 C_1}\right), \quad Z_2 = j\left(\frac{\omega^2 L_2 C_2 - 1}{\omega C_2}\right),$$

and

$$\frac{Z_1}{Z_2} = -\frac{\omega^2 C_2 L_1}{(1 - \omega^2 L_1 C_1)^2},$$

since, for the constant-k filter, $L_1C_1 = L_2C_2$. A filter of this type will pass all frequencies such that Z_1/Z_2 will lie between 0 and -4. Thus, $-\dfrac{\omega_c{}^2L_1C_2}{(1-\omega_c{}^2L_1C_1)^2} = -4$, the solution of which is

$$\omega_c = \frac{1}{4}\left(\sqrt{\frac{1}{L_2C_1}+\frac{16}{L_1C_1}} \pm \frac{1}{\sqrt{L_2C_1}}\right). \tag{92}$$

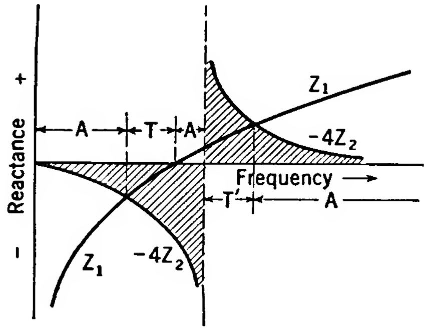


FIG. 40. Reactance curves for a non-confluent double band-pass filter, passing the two bands T and T'. This figure is *not* for the filter of Figs. 38 and 39.

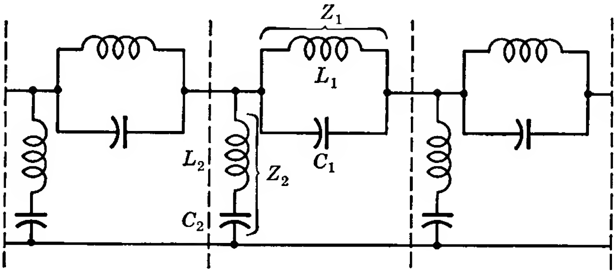


FIG. 41. A band-elimination filter of the constant-k type.

Two cutoff frequencies are obtained,

$$\left.\begin{aligned} f_c'' &= \frac{1}{8\pi}\left(\sqrt{\frac{1}{L_2C_1}+\frac{16}{L_1C_1}} + \frac{1}{\sqrt{L_2C_1}}\right) \\ \text{and} \qquad & \\ f_c' &= \frac{1}{8\pi}\left(\sqrt{\frac{1}{L_2C_1}+\frac{16}{L_1C_1}} - \frac{1}{\sqrt{L_2C_1}}\right). \end{aligned}\right\} \tag{93}$$

To find the upper and lower frequency limits, the impedance ratio is equated to zero. Thus, $Z_1/Z_2 = -\omega_c{}^2C_2L_1/(1-\omega_c{}^2L_1C_1) = 0$, and

from this, two values $\omega_c = 0$ and $\omega_c = \infty$ are obtained. From these two relations and from the preceding paragraph it follows that the band-elimination filter will pass all frequencies from zero to f_c' cycles per second, will attenuate all frequencies from f_c' to f_c'', and will pass all currents having frequencies between f_c'' and infinity. This filter then eliminates the band f_c' to f_c''.

The method of finding the design equations is similar to that used for the band-pass filter. If the two expressions of equation 93 are subtracted, it will be found that $(f_c'' - f_c') = \dfrac{1}{4\pi\sqrt{L_2C_1}}\cdot$ Since, for the constant-k filter, $L_1C_1 = L_2C_2$ and $Z_K = \sqrt{L_2/C_1} = \sqrt{L_1/C_2}$, this equation can be solved for C_1, and

$$C_1 = \frac{1}{4\pi(f_c'' - f_c')Z_K} \text{ farads.} \tag{94}$$

Infinite and zero values occur at $f_\infty = \dfrac{1}{2\pi\sqrt{L_1C_1}} = \dfrac{1}{2\pi\sqrt{L_2C_2}}\cdot$ Using these relations and the equations just obtained, it can easily be shown that

$$\left.\begin{aligned} L_2 &= \frac{Z_K}{4\pi(f_c'' - f_c')} \text{ henrys,} \quad L_1 = \frac{(f_c'' - f_c')Z_K}{\pi f_c' f_c''} \text{ henrys,} \\ \text{and} \quad & \\ C_2 &= \frac{(f_c'' - f_c')}{\pi f_c' f_c'' Z_K} \text{ farads.} \end{aligned}\right\} \tag{95}$$

Calculations for the various inductors and capacitors for a simple band-elimination filter are made from equations 94 and 95.

Composite Filters. The constant-k filters discussed in the preceding pages are satisfactory for some purposes, but the rigid requirements of modern communication have made necessary filters with better characteristics.

From Figs. 29 and 34, it is seen that the attenuation offered by constant-k filters increases gradually beyond the cutoff frequency. Filters are needed that offer very high attenuation close to the cutoff point. Thus, the *attenuation characteristic* of the constant-k filter is one factor limiting its use.

As shown by Figs. 25, 28, and 33, the iterative impedance of constant-k filters varies widely with frequency. It follows that a constant-k filter would offer a poor termination to transmission lines and telephone equipment. Thus, the *impedance characteristic* is another factor limiting its use.

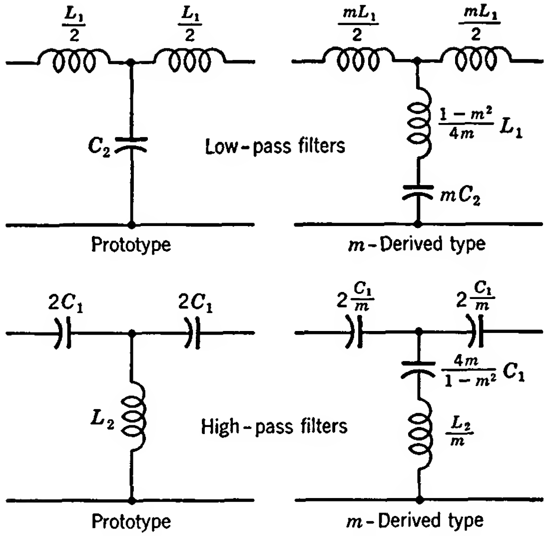


FIG. 42. Mid-series constant-k, or prototype, filter sections, and corresponding m-derived sections.

For these reasons, **composite filters,** rather than constant-k filters, are often used. These composite filters usually consist of constant-k sections terminated with m-derived half sections. Composite filters can be made to have very high attenuation close to the cutoff point, and to offer a uniform iterative impedance over the band of frequencies transmitted.

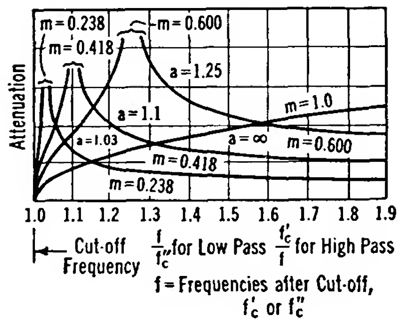


FIG. 43. Variations of attenuation with frequency ratios for m-derived and constant-k sections ($m = 1.0$).

***m*-Derived Sections.** Structures known as ***m*-derived sections** have good impedance characteristics and sharp cutoff frequencies. In Fig. 42 are shown T, or mid-series, sections, and as explained in the preceding section m-derived and constant-k sections are used together to make composite filters.

These m-derived sections have attenuation characteristics shown in Fig. 43. It will be noted that both m and a are shown on these curves, the relation between these two quantities being as follows

$$m = \sqrt{1 - 1/a^2}, \quad \text{and} \quad a = 1/\sqrt{1 - m^2}, \tag{96}$$

where

$$a = \frac{f_\infty}{f_c''} \text{ (L.P. filter), and } \frac{f_c'}{f_\infty} \text{ (H.P. filter).} \tag{97}$$

In this relation f_∞ is the frequency at which the attenuation of the m-derived section reaches a theoretical value of infinity (Fig. 43), and f_c is the cutoff frequency of the constant-k **prototype,** that is, the constant-k section to which the m-derived section is related by the factor m. Referring to Fig. 42, if $m = 1.0$, the derived sections become the constant-k prototype.

Impedance Relations in m-Derived Sections. The iterative impedance of a low-pass constant-k T section is given by equation **71,** and is

$$Z_{KT} = \sqrt{\frac{L_1}{C_2}}\sqrt{1 - \left(\frac{f}{f_c''}\right)^2} = Z_K\sqrt{1 - \left(\frac{f}{f_c''}\right)^2}, \tag{98}$$

where Z_K is the iterative impedance *at zero frequency* as given by equation 73. For the m-derived section of Fig. 42, equation 98 also applies. (See page 193 of reference 5.) The significance of this is that m and a do not appear in equation 98, and hence the midseries iterative impedance of an m-derived low-pass filter having any value of m or a is the same as the mid-series iterative impedance of the constant-k prototype. Thus, a low-pass m-derived T section can be connected to its constant-k prototype T section *without an impedance mismatch occurring at the junction* for any frequency.

As has been mentioned before, constant-k sections have poor iterative impedance characteristics, and m-derived sections are used to remedy this. But, the preceding paragraph brings out the point that the *midseries* iterative impedances for the prototype and the derived sections are the same. Thus, it becomes necessary to investigate the *midshunt* iterative impedance characteristics of π sections. The midshunt iterative impedance of a low-pass m-derived π section is (see page 209 of reference 5)

$$Z_{K\pi} = \frac{Z_K\left[1 - \left(\frac{f}{af_c''}\right)^2\right]}{\sqrt{1 - \left(\frac{f}{f_c''}\right)^2}}. \tag{99}$$

It is important to note that a, the ratio f_∞/f_c, enters this equation and to recall that a and m are related as shown in equation **96.** Thus, the *midshunt* iterative impedance of an m-derived low-pass π

filter section *varies with the value of a or m,* as shown in Fig. 44. From this figure it is seen that if $m = 0.6$ the midshunt iterative impedance (that is, the iterative impedance of an m-derived π section) is substantially constant from zero frequency to cutoff frequency.

It also can be proved that the cutoff frequencies of m-derived sections and their prototypes are identical. (See page 193 of reference 5.) It will be noted that Figs. 43 and 44 apply both to low-pass and high-pass filters.

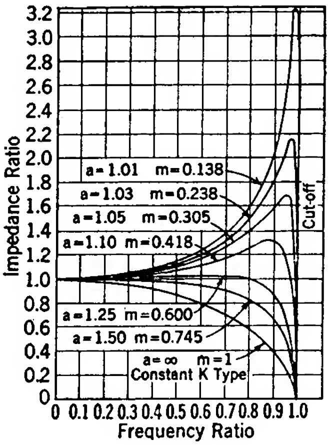


FIG. 44. Variations in the iterative impedance $Z_{K\pi}$ for various filter sections with respect to the impedance Z_K at $f = 0$ for the low-pass filter and $f = \infty$ for the high-pass filter. Frequency ratio f/f_c'' for low-pass filters, and f_c'/f for high-pass filters.

The characteristics of constant-k sections, m-derived sections, and their uses in composite filters, will now be summarized. (A) Constant-k sections have poor iterative-impedance characteristics as shown by Figs. 25, 28, and 33. (B) Constant-k sections do not have sharp cutoff characteristics as shown by Figs. 29 and 34, but they *do* have high attenuation at frequencies considerably beyond cutoff as indicated by these figures. (C) The midseries iterative impedance of an m-derived T section is the same, for any value of m, as the midseries iterative impedance of its constant-k prototype; hence, m-derived T sections of any value of m or a and constant-k prototype T sections can be connected together without reflection loss at the junction. (D) The midshunt iterative impedance of an m-derived π section is determined, at any frequency, by the values of m and a; hence, values of m and a may be selected such that very good iterative impedance characteristics can be obtained. (E) The cutoff frequencies of m-derived and constant-k prototype sections are the same; however, the m-derived section has theoretically infinite attenuation at a frequency determined by a and m.

Based on these facts, composite filters are constructed as follows: Sufficient constant-k T sections are used to give the desired attenuation considerably beyond cutoff; three such sections will provide sufficient attenuation for most purposes. From these prototypes an m-derived T section is designed, and this is then "split" into two **m-derived half sections,** which are used to terminate the constant-k sections so that

good impedance characteristics and sharp attenuation characteristics are obtained. The m-derived half sections are connected so that they offer midshunt terminations (which vary with m and can be made nearly independent of frequency) to networks external to the final composite filter. However, they offer midseries iterative impedances to constant-k T sections inside the filter, and hence internal reflections are prevented (Fig. 45).

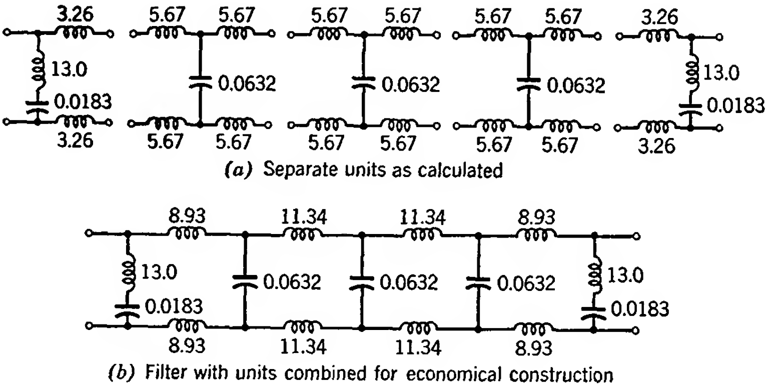


FIG. 45. Illustrating the design of a balanced low-pass composite filter. Inductances in millihenrys. Capacitances in microfarads.

Design of a Composite Filter. Suppose that a balanced low-pass composite filter cutting off at 8400 cycles, and with (theoretically) infinite attenuation at 10,300 cycles, is desired. This filter is to be used with lines and equipment having impedances of 600-ohms resistance. Three constant-k sections will provide sufficient attenuation considerably beyond the point of cutoff.

The m corresponding to an $a = 10{,}300/8400 = 1.225$ is, from equation 96, $m = 0.577$. From Fig. 44 it is seen that this m will give good impedance match, and hence *one* m-derived section can be used to obtain *both* the infinite attenuation and the good impedance match.

From equations 74 and 75 for the constant-k sections, $L_1 = 600/(3.1416 \times 8400) = 22.7$ millihenrys; also, $C_2 = 1/(3.1416 \times 8400 \times 600) = 0.0632$ microfarad. For the m-derived sections (see Fig. 42), $mL_1/2 = (0.577 \times 22.7)/2 = 6.55$ millihenrys. The value of the shunt capacitor is $mC_2 = 0.577 \times 0.0632 = 0.0365$ microfarad. Also, $(1 - m^2)L_1/4m = [(1 - 0.577^2) \times 22.7]/(4 \times 0.577) = 6.55$ millihenrys.

In the preceding paragraph an m-derived T section was designed.

Half of this is to be used to terminate each end of the constant-k sections. The complete filter will then consist of the sections of Fig. 45(a). The units would, for economy, be combined as shown in Fig. 45(b). In the final filter it is of interest to find that the 13.1-millihenry inductors and the 0.0183-microfarad capacitors shunting each end of the filter are tuned to a frequency of approximately 10,300 cycles, the frequency at which infinite attenuation is desired. The measured attenuation characteristics are shown in Fig. 46.

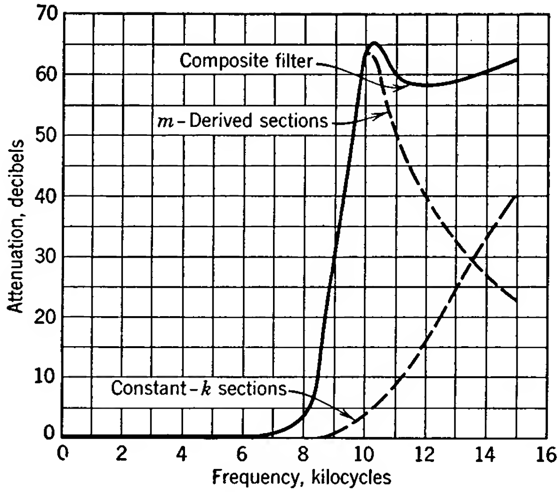


FIG. 46. Attenuation characteristics of the composite filter of Fig. 45, and estimated contribution made by each portion.

Suppose that a very sharp cutoff had been desired, with an infinite attenuation at about 9240 cycles, giving an $a = 9240/8400 = 1.1$. According to Fig. 44, an a of 1.1 or an m of 0.416 would not give a good termination. Therefore, one m-derived T section of $a = 1.1$ would be added internally to the composite filter of Fig. 45. Although in the previous discussions only a low-pass filter was designed, the procedure is the same for a high-pass composite filter.

Lattice-Type Filters.[5] The filters which have been previously considered were of the ladder type. Campbell developed[8] the lattice-type structure shown schematically in Fig. 11. With the lattice structure, sharper cutoff characteristics are obtainable and greater flexibility of design is possible. Campbell proved[9] that the lattice structure was as effective in filtering as any symmetrical structure of inductors and

capacitors can be made. Attention is directed to the fact that the configuration of the lattice structure is that of a Wheatstone bridge; for this reason such networks sometimes are referred to as **bridge structures.**

Quartz Crystal Filters.[7] In the preceding discussion of wave filters the inductors and capacitors were assumed lossless. Although this assumption holds closely for good capacitors, it does not hold for inductors. These losses cause the actual filter characteristics to differ from the ideal, especially with regard to the sharpness of the curves near the cutoff frequency. Also, the losses increase rapidly with frequency, and for this and other reasons it is difficult to construct filters using inductors and capacitors for frequencies above about 50,000 cycles.

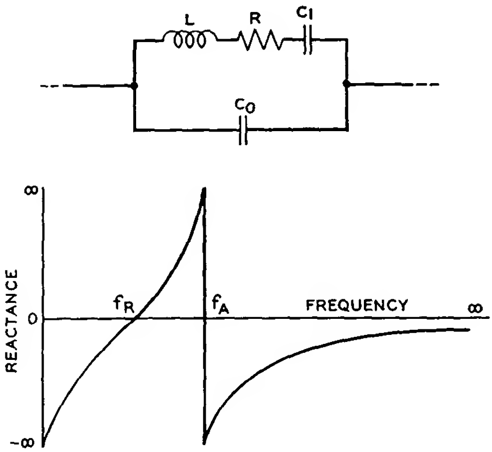


Fig. 47. A quartz crystal acts, when associated with a suitable electric circuit, as the network shown above. The reactance characteristics of this circuit are shown below. The resistance component is low at the resonant frequency f_R and high at the antiresonant frequency f_A. (Reference 11.)

This *was* one of the important reasons limiting the use of carrier telephone systems to the band just mentioned. Crystal filters, together with other important developments, have extended the upper limit to millions of cycles.

The use of quartz crystals was suggested[9] by Espenschied and has been made practicable by the work[10] of Mason and others. As is well known, crystals of quartz vibrate at certain frequencies but not at others. They are electrically equivalent to tuned circuits. The Q of a quartz crystal is as high as 20,000 or more.[11]

The equivalent electrical network and the reactance characteristics of a quartz crystal are shown in Fig. 47. The inductance L represents the mass reaction of the crystal against motion (inertia effect); the resistance R represents the energy dissipation within the crystal; the condenser C_1 represents the compliance[1] of the crystal; and C_0 is the capacitance of the circuit with the crystal not in motion. As indicated, there is one series resonant frequency f_R and one parallel or anti-

resonant frequency f_A of the crystal. For a given crystal, these have a ratio such that f_A is 0.4 per cent higher than f_R.[11] This ratio can be *reduced* and *controlled* by adding a capacitor in parallel with the crystal,[11] and if the capacitor has low losses, the circuit will retain a high ratio of reactance to resistance or Q. In Fig. 48 are shown a crystal filter of the ladder type, the corresponding reactance curves, and the resulting filter characteristics.

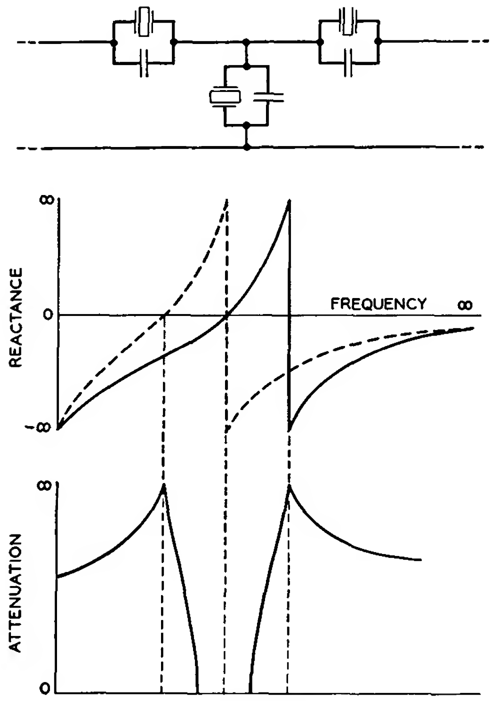


FIG. 48. Ladder-type filter network employing only crystals and capacitors, above; reactance characteristics, center (solid curve is series branch and dotted curve is shunt branch), attenuation characteristics, below. (Reference 11.)

Assume that the elements of Fig. 48 have the characteristics plotted in Fig. 47. For a single pass band, the series elements are resonant when the shunt unit is antiresonant. With this combination, maximum attenuation occurs when the shunt crystal is in resonance and when the series crystals are in antiresonance because of the minimum shunt impedance and the maximum series impedances at this frequency. When the crystals pass through the antiresonant frequency f_A, the resistance component of the impedance rises to a high value as in the usual parallel circuit.

As was previously explained, the point of antiresonance f_A is 0.4 per cent above the frequency of resonance f_R. Thus, the band passed is very narrow. For a 60,000-cycle carrier, the band would be only 240 cycles above and below the carrier, giving a total width of 480 cycles, a band too narrow for most purposes.[9] Thus, with the ladder-type crystal filter, the band of frequencies passed is always less than 0.8 per cent of the midfrequency value.

The filter of Fig. 48 is of an elementary type. Much progress has

been made in the design of crystal filters.[12, 13, 14] They are often made in the form of lattice structures giving improved characteristics. Sometimes the crystals are provided with divided electrodes[12, 13] that pass energy from one set of electrodes to the other only at frequencies at which the crystal vibrates. Natural quartz was used in crystal filters until about 1947, and since then synthetic crystals[15] have been used extensively.

REFERENCES

1. *American Standard Definitions of Electrical Terms.* A.I.E.E., 1941.
2. Ware, L. A., and H. R. Reed. *Communication Circuits.* John Wiley & Sons.
3. Everitt, W. L. *Communication Engineering.* McGraw-Hill Book Co.
4. Shea, T. E. *Transmission Networks and Wave Filters.* D. Van Nostrand Co.
5. Johnson, K. S. *Transmission Circuits for Telephonic Communication.* D. Van Nostrand Co.
6. Thiessen, Arthur E. Volume Control in Voice Circuit. *General Radio Experimenter,* June, 1931, Vol. 6, No. 1.
7. Mason, W. P. *Electromechanical Transducers and Wave Filters.* D. Van Nostrand Co.
8. Campbell, G. A. Physical Theory of the Electric Wave-Filter. *Bell System Tech. J.,* November, 1922, Vol. 1, No. 2.
9. Buckley, O. E. The Evolution of the Crystal Wave Filter. *Bell Telephone Quarterly,* January, 1937, Vol. 16, No. 1.
10. Mason, W. P. Electrical Wave Filters Employing Quartz Crystals as Elements. *Bell System Tech. J.,* July, 1934, Vol. 13, No. 3.
11. Mason, W. P. Quartz Crystal Filters. *Bell Lab. Record,* June, 1935, Vol. 13, No. 10.
12. Lane, C. E. Duplex Crystals. *Bell Lab. Record,* February, 1946, Vol. 24, No. 2.
13. Burns, G. K. Manufacture of Quartz Crystal Filters. *Bell System Tech. J.,* October, 1940, Vol. 19, No. 4.
14. Willis, E. S. A New Crystal Channel Filter for Broad Band Carrier Systems. *Elec. Eng.,* March, 1946, Vol. 65, No. 3.
15. Mason, W. P. New Low-Coefficient Synthetic Piezoelectric Crystals for Use in Filters and Oscillators. *Proc. I.R.E.,* October, 1947, Vol. 35, No. 10.

REVIEW QUESTIONS

1. Distinguish between an active and a passive network, and give an example of each.
2. What are meant by the terms rheostat, voltage divider, and potentiometer?
3. Strictly speaking, what is meant by the term load? How is it often used in communication?
4. Explain the difference between a midseries and a midshunt termination.
5. What is the difference between balanced and unbalanced networks?
6. What is the difference between symmetrical and unsymmetrical networks?
7. What is meant by the term equivalent networks? What is limited equivalence?

8. Which of the two network theorems accomplishes the same results in different ways? Where is each applied?
9. Give an application of the superposition theorem.
10. How would you measure the mutual impedance of a transformer?
11. How do iterative impedance and image impedances differ?
12. How do transition loss and reflection loss differ?
13. Explain what is meant by insertion loss. Give a practical example.
14. What is an electric wave filter? Name a few uses.
15. For a filter to pass a wave, what iterative impedance relations must hold?
16. How does the phase angle of the input impedance of a low-pass filter vary numerically with frequency? Repeat for a high-pass filter.
17. What is meant by saying that a filter is confluent?
18. What are composite filters, and why are they used?
19. What is meant by the term m-derived section?
20. Enumerate the steps in the design of a composite filter.
21. Explain how midseries and midshunt iterative impedances are of importance in composite filter design.
22. What is meant by the statement on page 187 regarding the Q of a quartz crystal?
23. Why are quartz crystals used in wave filters?
24. What is meant by saying that a crystal has divided electrodes?
25. What very important development in crystals occurred in about 1947?

PROBLEMS

1. In Fig. 4, Z_1 is a 0.05-microfarad capacitor, and Z_2 is a 0.12-henry inductor. Calculate the values to be used for midseries and midshunt terminations such as shown in Figs. 5(a) and 5(b).
2. Calculate the values for the equivalent T section and the equivalent π section of a network such as Fig. 4 in which $Z_1 = 0.15$ henry, and $Z_2 = 0.5$ microfarad.
3. In Fig. 9(a), $Z_A = 100$ ohms resistance, $Z_B = 50$ ohms resistance, and $Z_C = 100$ ohms resistance; $Z_D = 0.05$ microfarad, $Z_E = 0.05$ microfarad, and $Z_F = 0.01$ microfarad; $Z_7 = 450$ ohms resistance and 0.002 henry inductance, $E_g = 58$ volts open circuit at 1000 cycles, and $Z_L = 275$ ohms resistance. Calculate the magnitude and phase angle of the current through Z_L, the angle to be measured with respect to the generator voltage.
4. A transformer has the following characteristics $L_p = 0.159$ henry, $R_p = 193$ ohms, $M = 0.1585$ henry, secondary same as primary. For a frequency of 1000 cycles determine the equivalent unbalanced T and π networks. What is the coefficient of coupling? The turns ratio? The mutual impedance?
5. A generator has an internal impedance of 600 ohms resistance and an open-circuit voltage of 26.4 volts at 1000 cycles. It is connected to a load resistance of 768 ohms. Calculate the current through the load resistor using Thévenin's theorem and Norton's theorem.
6. Find the reflection loss for the circuit of Problem 5.
7. Design a pad to insert 7.5 decibels loss in a 600-ohm circuit.
8. Design a low-pass constant-k filter for a 600-ohm circuit and a cutoff at 5000 cycles.

9. Design a high-pass constant-k filter for a 600-ohm circuit and a cutoff at 100 cycles.
10. Design a composite filter for a 600-ohm circuit, for a cutoff frequency of 5000 cycles, and an $a = 1.2$.

CHAPTER *6*

TRANSMISSION LINES

Introduction. The preceding chapter considered the transmission of electric-signal energy through networks in which the elements were *lumped;* that is, existed as individual units of appreciable magnitude. This chapter will consider the transmission of electric-signal energy through open-wire lines in which the resistance, inductance, capacitance, and shunt conductance are uniformly *distributed,* rather than lumped.

An **open-wire circuit** is defined[1] as "a circuit made up of conductors separately supported on insulators." The open-wire telephone line is an example. Such a line is a **uniform line,** defined[1] as "a line that has identical electrical properties throughout its length." For convenience these will be called **transmission lines,** or merely **lines.**

These lines have **linear electrical constants,** defined[1] as "the series resistance, series inductance, shunt conductance, and shunt capacitance per unit length of line." Sometimes one overhead wire and earth return are used, constituting a **ground-return circuit.**[1] Only **metallic circuits,** in which the ground or earth forms no part,[1] will be considered in this chapter.

The range of frequencies transmitted for communication purposes over such lines is great. The upper limit of the frequencies used in telegraph systems is several hundred cycles. For voice-frequency telephone purposes, the band is from about 200 to 3500 cycles. Radio-broadcast program lines for amplitude-modulated broadcasting usually transmit a band from about 100 to 5000 cycles; for frequency-modulation, a band from about 30 to 15,000 cycles is sometimes used. Multichannel carrier-telephone circuits use frequencies up to 150,000 cycles. Radio frequencies up to, perhaps, one billion cycles are sometimes used (if only experimentally) on open-wire transmission lines.

The **power levels,** defined[1] as "an expression of the power being transmitted past any point in a system," of the signals on communication transmission lines cover a wide range of values. For telephone transmission the level is a few milliwatts or less. On radio transmission lines, such as are used to connect a radio transmitter to its sending antenna, the power level may be many kilowatts.

In the transmission of speech and programs, the distortion (page 85)

caused by the lines and associated equipment must be kept low. If appreciable *frequency distortion* exists, then the various frequency components of a complex speech or program signal wave will not be received in the same relative magnitude that they had at the sending end. If *non-linear distortion* is not minimized, then frequency components that were not present in the original signal will exist at the receiving end. *Delay distortion* will cause a shift in the relative phase positions of the various components because all frequency components are not transmitted with the same velocity. In telephony, some distortion can be tolerated, because reliability, intelligibility, and a certain degree of naturalness, rather than high quality, are the criteria of good service.

Transmission lines used in communication are **electrically long lines.** For a line to be electrically long, both the physical length of the line and the frequency of the signal wave being transmitted must be sufficiently great so that a considerable fraction of a wavelength exists on the line. Then, the instantaneous current at one point in the line is not the same as the instantaneous current at another point, and the instantaneous voltage between the wires at one point is not the same as that at another point. Simple electrical theory, such as is used with many electrical circuits, is not sufficient for studying communication transmission lines. Many 60-cycle power lines are classed as **electrically short lines,** because, so far as the fundamental frequency is concerned, they do not have an appreciable portion of a wavelength on them.

Power efficiency is of secondary consequence in telephone transmission because only a few milliwatts of power are contained in the signal wave and because line losses can be offset by the use of amplifiers. In a radio-frequency transmission line power efficiency may be of importance, because large amounts of power may be involved.

Linear Electrical Constants of a Transmission Line. In studying transmission lines it must be kept clearly in mind that the line is more than just two parallel wires; actually, it is an electric network composed of many sections.

Series Resistance R. The series resistance R is a "constant" that is found by dividing the total power losses occurring in the *series* portion of unit length of transmission line by the square of the current in that section. Thus, the series resistance losses are *losses caused by the current flow and the accompanying magnetic field,* including the direct-current resistance of the wire, and the energy losses, if any, caused by skin effect, eddy currents, and hysteresis. Series resistance R is usually expressed in ohms per *loop mile* for telegraph and tele-

phone lines, and in ohms or microhms *per loop meter* for radio transmission lines. A **loop mile** consists of *two* miles of wire.

Series Inductance L. Series inductance L is the series self-inductance of the line wires. The current in a line wire causes an alternating magnetic field around the line wire, and this field induces an electromotive force in series in the line wire. This back electromotive force is directly proportional to the inductance. Because the currents in the two line wires are in opposite directions, the magnetic effects tend to cancel, and, if the two wires of a transmission line are close together, the self-inductance is low. Series self-inductance is usually expressed in henrys per *loop mile* for telegraph and telephone circuits, and in millihenrys or microhenrys per *loop meter* for radio transmission lines.

Shunt Conductance G. Shunt conduction G is a "constant" that is found by dividing the total power losses occurring in the *parallel* portion of unit length of line by the square of the voltage between the length of line ($P = E^2/R = E^2G$). Shunt conductance losses are *losses that are caused by the voltage existing between line wires.* These losses are somewhat involved and include losses due to leakage currents, dielectric hysteresis losses, etc. For telephone transmission lines, the shunt conductance, or **leakage conductance,** as it is sometimes called, is expressed in micromhos per *mile of two-wire line.* In radio transmission lines the shunt conductance is usually neglected.

Shunt Capacitance C. Shunt capacitance C exists between the two wires of a transmission circuit, and an alternating current will flow through this capacitance if an alternating voltage exists between the line wires. For telephone transmission lines the capacitance is usually expressed in microfarads per *mile of two-wire line.* For radio transmission lines, the unit is usually microfarads or micromicrofarads *per meter* of line.

Electromagnetic Waves along Wires. Electric energy is transmitted along wires by electromagnetic waves. These consist of variations in the magnitude and direction of the electric and magnetic fields produced by the line voltages and currents. These two fields, as they travel together, transfer the electric energy from one point to another and constitute an **electromagnetic wave.**

Electric energy cannot be transferred unless an electric and a magnetic field exist simultaneously. If these two fields are in time phase, the power is directly proportional to their product; if out of phase, it is proportional to their product and the cosine of the time phase angle

between them. The maximum velocity with which an electromagnetic wave can travel is that of light, or approximately 186,300 miles, or 300,000,000 meters, per second. This applies to waves in free space; waves along wires travel more slowly.

In Fig. 1 is shown a portion of an infinitely long line connected to a source of alternating voltage. Current and voltage impulses will be distributed along the line somewhat as indicated. The dots (heads of arrows) and the crosses (tails of arrows) represent the magnetic field around the wires, and the arrows between wires represent the electric

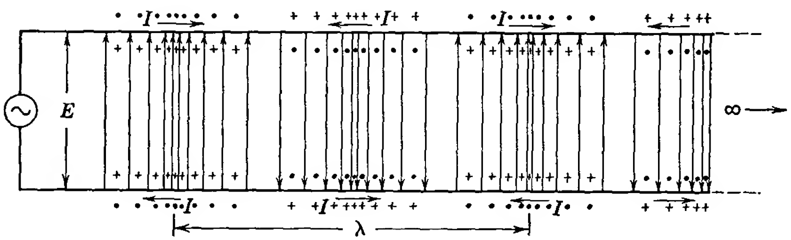


FIG. 1. The distribution of current and the accompanying magnetic field are shown by the dots and crosses adjacent to the wires. The distribution of the voltage and the accompanying electric field are shown by the arrows between wires.

field. The magnitudes of the current and the voltage, and the corresponding intensity of the fields, are indicated by the relative number of arrows, dots, and crosses. The wavelength λ is the distance between any two corresponding values, and the velocity of propagation V is the distance traveled per second. The wavelength is

$$\lambda = V/f, \tag{1}$$

since, at a frequency of f cycles per second, f complete waves are produced on the line per second, and these can travel only a distance of V miles during a second. An end view of the magnetic and the electric lines of force constituting an electromagnetic wave along two parallel wires is shown in Fig. 2.

Derivation of Transmission Equations. An open-wire line (or a cable) is composed of a large number of elements dl, as shown in Fig. 3. If current flows through an impedance, the voltage drop is Iz, and therefore the voltage drop along the element of line dl will be

$$dE = -Izdl \quad \text{or} \quad \frac{dE}{dl} = -Iz. \tag{2}$$

As a current progresses down the line, a certain amount is lost through parallel admittance y consisting of the shunt conductance and the ca-

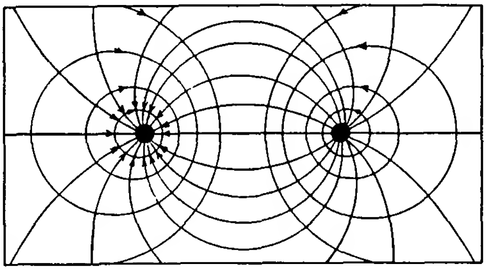


FIG. 2. End view of the line of Fig. 1, showing the electric field between the two wires, and the magnetic field around the wires.

pacitance of each element. This current loss at each element is Ey, and thus

$$dI = -Eydl \quad \text{or} \quad \frac{dI}{dl} = -Ey. \tag{3}$$

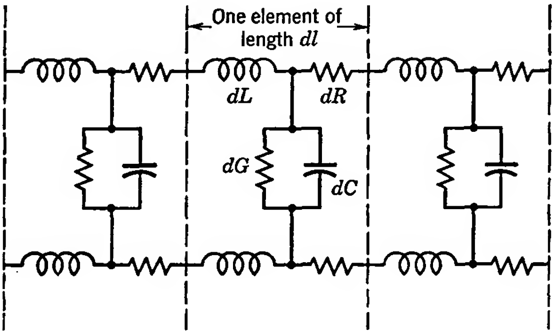


FIG. 3. A transmission line consists of distributed resistance, inductance, capacitance, and conductance.

It is now necessary to differentiate these two equations to eliminate E and I. Thus, when equation 2 is differentiated with respect to l it becomes

$$\frac{d^2E}{dl^2} = -z\frac{dI}{dl}; \tag{4}$$

similarly, equation 3 becomes

$$\frac{d^2I}{dl^2} = -y\frac{dE}{dl}. \tag{5}$$

The undesired variables can now be eliminated by substituting in equations 4 and 5 the values from equations 3 and 2 respectively. Then,

$$\frac{d^2E}{dl^2} = zyE, \tag{6}$$

and

$$\frac{d^2I}{dl^2} = zyI. \tag{7}$$

These two equations are simple linear differential equations of the second order, and a knowledge of differential equations is required for their solution. Without presenting the details, the general solution for equations of this form is

$$E = A_1\epsilon^{l\sqrt{zy}} + B_1\epsilon^{-l\sqrt{zy}} \tag{8}$$

and

$$I = A_2\epsilon^{l\sqrt{zy}} + B_2\,\epsilon^{-l\sqrt{zy}}. \tag{9}$$

The four constants of integration A_1, B_1, A_2, and B_2 must now be found. To do this, equations 8 and 9 are first differentiated with respect to l. They then become

$$\frac{dE}{dl} = A_1\sqrt{zy}\,\epsilon^{l\sqrt{zy}} - B_1\sqrt{zy}\,\epsilon^{-l\sqrt{zy}} \tag{10}$$

and

$$\frac{dI}{dl} = A_2\sqrt{zy}\,\epsilon^{l\sqrt{zy}} - B_2\sqrt{zy}\,\epsilon^{-l\sqrt{zy}}. \tag{11}$$

It will be noted that these two expressions are equal to equations 2 and 3, respectively. It can therefore be written that at the sending end, where $l = 0$,

$$I_sz = -\,(A_1\sqrt{zy} - B_1\sqrt{zy}) \tag{12}$$

and

$$E_sy = -\,(A_2\sqrt{zy} - B_2\sqrt{zy}). \tag{13}$$

These values are preceded by a negative sign because the current I and the voltage E in any part of the line are less than the sending-end values I_s and E_s in the infinite line (having no reflection) under consideration.

If l is also assumed equal to zero in equations 8 and 9, then E becomes the sending-end voltage E_s, and

$$E_s = A_1 + B_1, \; A_1 = E_s - B_1, \quad \text{and} \quad B_1 = E_s - A_1. \tag{14}$$

Similarly,

$$I_s = A_2 + B_2, \; A_2 = I_s - B_2, \quad \text{and} \quad B_2 = I_s - A_2. \tag{15}$$

When these values of A_1 and B_1 are substituted in equations 12 and 13,

$$A_1 = E_s/2 - I_s/(2\sqrt{y/z}) \quad \text{and} \quad B_1 = E_s/2 + I_s/(2\sqrt{y/z}). \tag{16}$$

Also,

$$A_2 = I_s/2 - E_s\sqrt{y/z}/2 \quad \text{and} \quad B_2 = I_s/2 + E_s\sqrt{y/z}/2. \tag{17}$$

Furthermore, when these values of A_1, B_1, A_2, and B_2 are substituted in equations 8 and 9,

$$E = E_s\left(\frac{\epsilon^{l\sqrt{zy}} + \epsilon^{-l\sqrt{zy}}}{2}\right) - I_s\sqrt{z/y}\left(\frac{\epsilon^{l\sqrt{zy}} - \epsilon^{-l\sqrt{zy}}}{2}\right) \tag{18}$$

and

$$I = I_s\left(\frac{\epsilon^{l\sqrt{zy}} + \epsilon^{-l\sqrt{zy}}}{2}\right) - E_s\sqrt{y/z}\left(\frac{\epsilon^{l\sqrt{zy}} - \epsilon^{-l\sqrt{zy}}}{2}\right). \tag{19}$$

Although it would be possible to use these equations in this form, they can be simplified by the use of hyperbolic functions. In textbooks on this subject it is shown that

$$\frac{\epsilon^{x} + \epsilon^{-x}}{2} = \cosh x \quad \text{and} \quad \frac{\epsilon^{x} - \epsilon^{-x}}{2} = \sinh x, \tag{20}$$

where $\cosh x$ and $\sinh x$ are the hyperbolic cosine and sine, respectively. Accordingly, equations 18 and 19 become

$$E = E_s \cosh l\sqrt{zy} - I_s\sqrt{z/y} \sinh l\sqrt{zy} \tag{21}$$

and

$$I = I_s \cosh l\sqrt{zy} - E_s\sqrt{y/z} \sinh l\sqrt{zy}. \tag{22}$$

These two equations give the voltage E or the current I (in either maximum or effective values) at any point l miles from the sending end of a line having a series impedance of $z = R + jX$ (where $X = 2\pi fL$)

ohms per loop mile, and a parallel admittance of $y = G + jB$ (where $B = 2\pi fC$) mhos per loop mile (page 194).

If the values in equations 12 and 13 had been preceded by a positive instead of a negative sign and if E_r and I_r had been used, equations for voltage and current at any point in the line in terms of the receiving-end voltage and current and the distance l from the receiving end would have been obtained. These equations are

$$E = E_r \cosh l\sqrt{zy} + I_r\sqrt{z/y} \sinh l\sqrt{zy} \tag{23}$$

and

$$I = I_r \cosh l\sqrt{zy} + E_r\sqrt{y/z} \sinh l\sqrt{zy}. \tag{24}$$

Equations 21, 22, 23, and 24 are general transmission equations applying, for steady-state conditions, to telephone lines and cables. Other methods of deriving these equations are also available (references 2 to 7 inclusive).

Electromagnetic Wave Propagation. Suppose that for the infinite line of the preceding section a point an infinite distance from the sending end is considered. At this point, $l = \infty$ for equations 21 and 22, and these can be equated to zero and become

$$E_s \cosh l\sqrt{zy} = I_s\sqrt{z/y} \sinh l\sqrt{zy} \tag{25}$$

and

$$I_s \cosh l\sqrt{zy} = E_s\sqrt{y/z} \sinh l\sqrt{zy}. \tag{26}$$

These equations can be written

$$E_s = I_s\sqrt{\frac{z}{y}}\,\frac{\sinh l\sqrt{zy}}{\cosh l\sqrt{zy}} = I_s\sqrt{z/y} \tanh l\sqrt{zy} \tag{27}$$

and

$$I_s = E_s\sqrt{\frac{y}{z}}\,\frac{\sinh l\sqrt{zy}}{\cosh l\sqrt{zy}} = E_s\sqrt{y/z} \tanh l\sqrt{zy}, \tag{28}$$

since from the theory of hyperbolic functions

$$\sinh x/\cosh x = \tanh x.$$

An examination of a table of hyperbolic functions will show that, when $l = \infty$, $\tanh l\sqrt{zy} = 1.0$; therefore, equations 27 and 28 become

$$E_s = I_s\sqrt{z/y} = I_sZ_s \tag{29}$$

and

$$I_s = E_s\sqrt{y/z} = E_s/Z_s, \tag{30}$$

where Z_s is the **sending-end impedance** of the infinitely long line considered, measured between the two line input terminals. If equations 29 and 30 are substituted in the general expressions 22 and 21, respectively, then

$$E = E_s(\cosh l\sqrt{zy} - \sinh l\sqrt{zy}) \tag{31}$$

and

$$I = I_s(\cosh l\sqrt{zy} - \sinh l\sqrt{zy}). \tag{32}$$

Since $\epsilon^{-x} = (\cosh x - \sinh x)$, these equations become

$$E = E_s\epsilon^{-l\sqrt{zy}} \quad \text{or} \quad E = E_s\epsilon^{-l\gamma} \tag{33}$$

and

$$I = I_s\epsilon^{-l\sqrt{zy}} \quad \text{or} \quad I = I_s\epsilon^{-l\gamma}, \tag{34}$$

where

$$\gamma = \sqrt{zy} \tag{35}$$

and is called the **propagation constant** per loop mile for the line under consideration. Like equations 21 and 22, these equations also give the voltage or current at any point along the line in terms of the sending-end values.

Propagation Constant. In the preceding section the propagation constant per mile was shown to be $\gamma = \sqrt{zy}$ This expression can be expanded by substituting the values of z and y given on page 198. That is,

$$\gamma = \sqrt{(R + jX)\ (G + jB)}. \tag{36}$$

It is evident that equation 36 contains both real and imaginary terms. Thus, γ must also contain real and imaginary components, and therefore it can be written that

$$\gamma = \alpha + j\beta = \sqrt{(R + jX)\ (G + jB)}$$
$$= \sqrt{(R + j\omega L)\ (G + j\omega C)}. \tag{37}$$

If equation 37 is squared,

$$(\alpha + j\beta)^2 = (R + jX)\ (G + jB) \tag{38}$$

and

$$\alpha^2 - \beta^2 + 2j\alpha\beta = (RG - XB) + j(GX + BR). \tag{39}$$

If the real terms and the imaginary terms are equated separately, then

$$\alpha^2 - \beta^2 = RG - XB \tag{40}$$

and

$$2\alpha\beta = GX + BR. \tag{41}$$

When equation 41 is solved for α, it becomes

$$\alpha = \frac{GX + BR}{2\beta}; \tag{42}$$

and, when it is substituted in equation 40, this equation equals

$$\frac{(GX + BR)^2}{4\beta^2} - \beta^2 = RG - XB \tag{43}$$

or

$$\frac{(GX + BR)^2}{4} - \beta^4 = (RG - XB)\beta^2, \tag{44}$$

and

$$\beta^4 + \beta^2(RG - XB) - \frac{(GX + BR)^2}{4} = 0. \tag{45}$$

From this equation β can be found and becomes

$$\beta = \sqrt{\tfrac{1}{2}\left(\sqrt{(R^2 + X^2)\,(G^2 + B^2)} - (RG - XB)\right)}. \tag{46}$$

Similarly,

$$\alpha = \sqrt{\tfrac{1}{2}\left(\sqrt{(R^2 + X^2)\,(G^2 + B^2)} + (RG - XB)\right)}. \tag{47}$$

The values of X and B can be replaced with the fundamental line constants, and these two equations then become

$$\beta = \sqrt{\tfrac{1}{2}\left(\sqrt{(R^2 + \omega^2L^2)\,(G^2 + \omega^2C^2)} - (GR - \omega^2LC)\right)} \tag{48}$$

and

$$\alpha = \sqrt{\tfrac{1}{2}\left(\sqrt{(R^2 + \omega^2L^2)\,(G^2 + \omega^2C^2)} + (GR - \omega^2LC)\right)}. \tag{49}$$

The values of β and α are in units corresponding to the units of the fundamental constants. If R, L, G, and C are expressed per loop mile, then β is the **phase constant** (formerly called wavelength constant) per loop mile, and α is the **attenuation constant** per loop mile.

The propagation constant per unit length of a uniform line is[1] the "natural logarithm of the ratio of the current at a point of the line to the current at a second point, at unit distance from the first point along

the line in the direction of transmission, when the line is infinite in length, or is terminated in its characteristic impedance." The attenuation constant is[1] "the *real* part of the propagation constant," and the phase constant is "the *imaginary* part of the propagation constant."

Characteristic Impedance. It was shown by equation 30 that for an infinite uniform line $I_s = E_s/\sqrt{z/y}$, where I_s is the sending-end current, E_s is the sending-end voltage, z is the linear series line impedance, and y is the linear parallel, or shunt, line admittance (page 198). From this relation,

$$\frac{E_s}{I_s} = \sqrt{\frac{z}{y}} = \sqrt{\frac{R + jX}{G + jB}} = \sqrt{\frac{R + j\omega L}{G + j\omega C}} = Z_O, \tag{50}$$

where Z_O is the **characteristic impedance,** defined[1] as "the ratio of an applied potential difference to the resultant current at the point where the potential difference is applied, when the line is of infinite length." The term characteristic impedance is applied correctly only to uniform lines with distributed constants (page 192).

When an alternating voltage is applied to a uniform infinite line of length L_∞, electromagnetic waves are propagated down the line toward infinity and no energy is reflected back to the sending end. The input impedance Z_i *measured* (with an impedance bridge) on this line of length L_∞, gives the characteristic impedance Z_O. This value is the same for all identical lines. Suppose that some finite length, such as 200 miles, is removed from the infinite line L_∞, and that the input impedance Z_i' of the remainder of the line ($L_\infty - 200$) is measured. This impedance Z_i' must equal Z_i because removing the 200 miles does not appreciably alter the infinite line. But the portion of the line ($L_\infty - 200$) of impedance Z_i' really acts as a termination or load to the 200-mile section. It terminates it without reflection, the wave entering it without encountering an impedance discontinuity. Since both the impedances Z_i and Z_i' equal the characteristic impedance Z_O as previously defined, it follows that, when a finite line is terminated with an impedance load equal to its characteristic impedance, then the electromagnetic wave received at the end of the finite line will enter the termination without reflection just as it entered the portion of the infinite line ($L_\infty - 200$). Hence the characteristic impedance Z_O of a finite line is equal to the input impedance of the line when it is terminated in its characteristic impedance Z_O.

The characteristic impedance of a line is that value of impedance which will terminate a finite length of line so that no wave reflection

will occur at the distant end. It should be noted from equation 50 that the characteristic impedance varies with frequency.

If it is desired to know the characteristic impedance of a line (pages 221 and 222), this value either can be computed from equation 50 or can be measured. In order to measure the characteristic impedance of a line, the impedance Z_{oc} is first measured with the distant end of the line open, and then Z_{sc} is measured with the distant end short-circuited. Then, from equations 60 and 64, given on page 207, and the relation $\tanh x = \dfrac{1}{\coth x}$,

$$Z_O = \sqrt{Z_{oc}Z_{sc}} = \sqrt{(\sqrt{z/y}\coth l\gamma)\,(\sqrt{z/y}\tanh l\gamma)} = \sqrt{z/y}. \qquad (51)$$

If the resistance and shunt conductance are assumed negligible, the characteristic impedance given by equation 50 becomes

$$Z_O = \sqrt{L/C} \qquad (51a)$$

and is sometimes called the **natural** or **surge impedance** of the line.

The Significance of Z_O, α, and β. From equations 30 and 50 it is seen that the sending-end current I_s equals the sending-end voltage E_s divided by the characteristic impedance Z_O if the line extends to infinity or is terminated in its characteristic impedance Z_O. For telephone lines operating at audio frequencies, Z_O has a small leading angle, and the sending-end, or input, current will lead the impressed sending-end voltage by this angle.

As this input current progresses down the line, some of the current is lost at the shunt conductance dG and shunt capacitance dC of Fig. 3. The rate at which the current is weakened, or attenuated, as it progresses down the line is determined by the attenuation constant; that is, by the real part of the propagation constant. Hence, from equations 34 and 37,

$$I_r = I_s\epsilon^{-l\alpha} \quad \text{or} \quad I_r = I_s 10^{-0.05ln}, \qquad (52)$$

where l is the line length, α is in nepers found by equations 37 or 49, and n is in decibels (page 86). Because the current is attenuated the accompanying magnetic field component of the electromagnetic wave also is attenuated.

As the current flows toward the distant end, voltage drops occur along the line wires because of the series resistance dR and the series inductance dL of Fig. 3. The rate at which the voltage is attenuated is determined by the attenuation constant α, and from equations 33 and 37,

$$E_r = E_s\epsilon^{-l\alpha} \quad \text{or} \quad E_r = E_s 10^{-0.05ln} \qquad (53)$$

where l, x, and n are as explained for equation 52. Because the voltage is attenuated, the accompanying electric field component of the electromagnetic signal wave also is attenuated.

Assume that an alternating signal voltage has just been impressed at the sending end of a line; for instance, at the left of Fig. 3. Before a voltage can exist at the end of a line section, the distributed capacitance of the section must be charged; this requires time, because the charging-current flow is limited by the series resistance and the inductance. As the voltage builds up across one section, the distributed capacitance of the next section starts to charge, and in this way the wave is propagated along the line. Time is, therefore, required for an electromagnetic wave to flow along a line. Thus the *instantaneous magnitudes* of the currents will not be the same at the various points along a line. The current may be maximum at one point, and zero at another point at the same instant. There is a *time phase angle* between current at one point and current at another point along the line. Likewise, there will be a *time phase angle* between the voltage at one point, and the voltage at another point along the line. These phase angles between *current* at one point and *current* at another point, or between *voltage* at one point and *voltage* at another point, are determined by the phase constant β.

The value of β (in radians) can be found by equations 37 or 48. The value of radians (when multiplied by 57.3°) determines the difference in phase in degrees. Corresponding points of the electromagnetic wave are 360° apart, and, as Fig. 4 indicates, 360° is one **wavelength.** Since β is the phase shift per unit length of line, the length of line required to shift the wave 360° or the length of one wavelength is

$$\lambda = \frac{2\pi}{\beta}. \tag{54}$$

It is known, however, that $\lambda = V/f$, where V is the wave velocity at the frequency f. Thus,

$$\frac{2\pi}{\beta} = \frac{V}{f} \quad \text{and} \quad V = \frac{2\pi f}{\beta} = \frac{\omega}{\beta}. \tag{55}$$

Input Impedance of Any Line with Any Termination. As explained on page 202, the input impedance of a line terminated in its characteristic impedance Z_O equals its characteristic impedance Z_O. Telephone lines are usually terminated in impedances closely approximating their characteristic impedances so that the maximum signal energy will be extracted from the oncoming wave and little reflection

will occur. Sometimes it is desirable to terminate transmission lines in impedances other than their characteristic impedances, and, if this is done, the input, or **driving-point, impedance,** *will not* equal the terminating impedance.

If a line is terminated with some impedance Z_r that *does not* equal Z_O, then an impedance mismatch exists, all the energy in the wave *is not* absorbed, and a wave is reflected back toward the sending end.

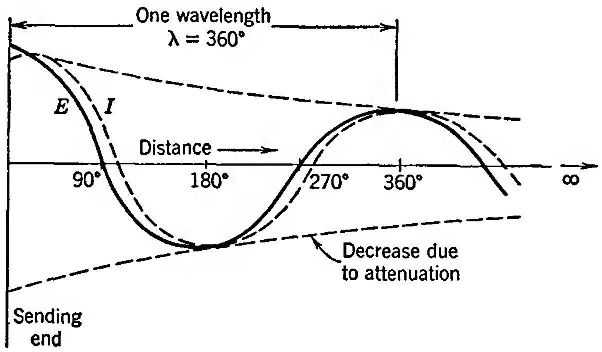


FIG. 4. The instantaneous voltage and current distribution along an infinite transmission line having loss is as shown. The current values are "ahead" of the voltage values because the characteristic impedance has a negative angle; that is, the reactive component is capacitive. Note that the X-axis values are *distance* along the line, plotted in degrees.

The wave that arrives back at the sending end combines with the wave entering the line from the signal source (generator) connected at the sending end. The exact manner of combining depends on the phase relations between the two components. The input impedance when the line termination *is not* Z_O will equal the sending-end voltage E_s divided by the sending-end current I_s. The sending-end voltage E_s will be the resultant of an **initial voltage wave** entering the line and a **reflected voltage wave** arriving back from the mismatched impedance Z_r. The sending-current I_s will be the resultant of an **initial current wave** and a **reflected current wave.**

In deriving the equation for the input impedance of a line terminated in any impedance Z_r, instead of its characteristic impedance Z_O, it is assumed first that the line is terminated in Z_O, and hence the magnitudes of the initial voltage and current waves can be found. The portions of these waves that reach the receiving end are determined by the length of line and the attenuation constant. The magnitudes and the phase relations of the reflected components depend on the magnitude

and the angle of the terminating impedance Z_r. Also, the magnitudes and phase relations of the reflected components that arrive back at the sending end and combine with the initial components to determine E_s and I_s depend on the propagation constant and the length of line.[6, 8]

The equation for the input impedance of a line with *any termination* Z_r can be found from equations 23 and 24. The line under consideration is assumed to be l miles in length, to have a propagation constant of $\gamma = \sqrt{zy}$ per mile, and to have a characteristic impedance of $Z_O = \sqrt{z/y}$. Then, from equation 23 the sending-end voltage E_s will be

$$E_s = E_r \cosh l\gamma + I_r Z_O \sinh l\gamma, \tag{56}$$

and from equation 24 the sending-end current will be

$$I_s = I_r \cosh l\gamma + E_r/Z_O \sinh l\gamma. \tag{57}$$

The voltage E_r at the receiving end is $E_r = I_r Z_r$, and, when this is substituted in equations 56 and 57, and when equation 56 is divided by equation 57 to obtain the input impedance,

$$Z_s = \frac{E_s}{I_s} = Z_O \left(\frac{Z_r \cosh l\gamma + Z_O \sinh l\gamma}{Z_O \cosh l\gamma + Z_r \sinh l\gamma} \right). \tag{58}$$

All terms in this equation are complex numbers, and the input, or driving-point, impedance Z_s measured at the sending end also is a complex number.

Input Impedance of Any Open-Circuited Line. The driving-point, or input, impedance of any such line or cable can be found by letting Z_r of equation 58 equal infinity. Or, it can be found as will be explained. When the distant end of a line is open the current at that point must be zero; thus, equation 23 becomes

$$E_{oc} = E_r \cosh l\sqrt{zy}, \tag{59}$$

and equation 24 equals

$$I_{oc} = E_r \sqrt{y/z} \sinh l\sqrt{zy}.$$

The sending-end impedance of the open-circuited line will be

$$Z_{oc} = \frac{E_{oc}}{I_{oc}} = \frac{E_r \cosh l\sqrt{zy}}{E_r\sqrt{y/z}\sinh l\sqrt{zy}} = \sqrt{z/y}\coth l\gamma, \tag{60}$$

since $\dfrac{\cosh x}{\sinh x} = \coth x$, and by equation 35 $\gamma = \sqrt{zy}$. Equation 60 can also be written

$$Z_{oc} = Z_O \coth l\gamma, \tag{61}$$

because from equation 50 the characteristic impedance $Z_O = \sqrt{z/y}$.

Input Impedance of Any Short-Circuited Line. The driving-point, or input, impedance of any such line or cable can be found by letting Z_r of equation 58 be zero, or by the following method. No voltage can exist across a short circuit, and thus, for the short-circuited condition, $E_r = 0$ in equations 23 and 24. It follows that

$$E_{sc} = I_r\sqrt{z/y}\sinh l\sqrt{zy} \tag{62}$$

and

$$I_{sc} = I_r \cosh l\sqrt{zy}. \tag{63}$$

The sending-end impedance of a short-circuited line will then be

$$Z_{sc} = \frac{E_{sc}}{I_{sc}} = \frac{I_r\sqrt{z/y}\sinh l\sqrt{zy}}{I_r\cosh l\sqrt{zy}} = \sqrt{z/y}\tanh l\gamma, \tag{64}$$

since $\dfrac{\sinh x}{\cosh x} = \tanh x$, and, by equation 35, $\gamma = \sqrt{zy}$. Equation 64 can also be written

$$Z_{sc} = Z_O \tanh l\gamma, \tag{65}$$

because, from equation 50, the characteristic impedance $Z_O = \sqrt{z/y}$.

Wave Reflection on Lines (and Cables). If an open-wire transmission line or a cable is not terminated in its characteristic impedance Z_O, some of the energy that reaches the distant end will be reflected. The reflected energy, in the form of a **reflected wave,** travels back toward the sending end. As it travels, it interferes with the oncoming **initial wave** that the source of signal energy is impressing at the sending end.

The reflection and interference phenomena that occur can be studied mathematically or graphically.[6] In the following discussions it will be assumed that the lines (or cables) are *lossless.* The lines will be either open circuited or short circuited at the distant end.

Voltage Distribution on an Open-Circuited Lossless Line. Figure 5 shows a generator of open-circuit voltage E_g and internal impedance $Z_g = Z_O$ connected to a line that is five-fourths wavelength long; that is, $V/f = \lambda$, and $l/\lambda = 5/4$, where l is the line length. The condition shown in Fig. 5 is *at the instant that the generated alternating signal voltage E_g is a maximum positive value.*

At this instant, the distribution of the electric field and the *initial voltage wave* along the line are as shown by the solid arrows. Because of the finite velocity of propagation, at a distance $\frac{1}{2}\lambda$ from the sending

end, and *at the instant under consideration,* the initial voltage wave traveling toward the distant end will be directed as indicated. Since $\frac{1}{2}\lambda$ equals 180°, the initial wave component at a distance $\frac{1}{2}\lambda$ from the sending end is equal *and opposite* to the initial wave component at the sending end. This component was created by the voltage when the impressed voltage was at the position $-180°$. At a distance λ from the sending end the initial voltage wave component will be directed downward, because it represents an impulse sent into the line when the impressed signal voltage was at the position marked $-360°$.

It will be noted that *at the instant under consideration* (when the impressed voltage is a maximum positive value) the initial impulse

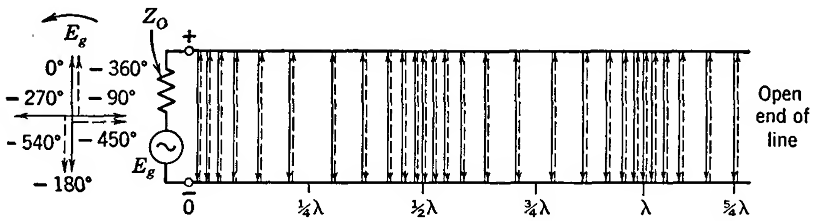


FIG. 5. At the instant that the generator voltage E_g is at the maximum positive value the initial voltage wave (solid arrows) and reflected voltage wave (broken arrows) are as shown. In this figure the intensities of the voltage and the electric field are shown by the density of the arrows. In Fig. 6 intensities are indicated to simplify drawing by distances from the X axis.

arriving at the distant end is zero. This is because the voltage at $-450°$ ($\frac{5}{4}\lambda = 450°$) was passing through zero. The *reflected wave component* at point λ on the line has traveled a distance $\frac{5}{4}\lambda$ or 450° from the generator to the end of the line, was reflected at the end of the line, and has traveled back $\frac{1}{4}\lambda$ or 90° to point λ. This reflected wave component (shown by broken arrows) was created at $-540°$ as shown by the broken arrow at $-540°$.

The preceding discussion assumes that the voltage wave is reflected at the open-circuited end of the line without a change in phase. This is true, because electric lines of force are assumed to extend from a positive wire to a negative wire, and, since there is no conducting path at the distant end, no interchange of charges from wire to wire can occur. The instantaneous directions of action (and not of travel) of the reflected-voltage component and the initial-voltage component are the same, and, for this reason, *the electric field and the voltage component of an electromagnetic wave are reflected with no change in phase from an open circuit at the end of a line.*

The preceding discussion and Fig. 5 considered the voltage distribution at the instant that the impressed voltage was a maximum positive value. This also is considered in Fig. 6(*a*). Both illustrations indicate that the *resultant voltage wave at the instant under consideration is zero at all points along the line.* This resultant voltage wave is the sum of the instantaneous initial and reflected waves.

The distributions of the initial, reflected, and resultant voltage waves at other positions of the instantaneous applied voltage are also shown in Fig. 6. It will be noted that in Fig. 6(*a*) the resultant voltage is zero, that in Fig. 6(*b*) the initial and reflected waves add directly, that in Fig. 6(*c*) the resultant again is zero, and that in Fig. 6(*d*) the waves again add directly. If voltmeters that indicate effective values of voltage (and not instantaneous values) were connected across the line at regular intervals, a plot of the instrument readings (effective values) would be as indicated by Fig. 6(*e*). Such a plot of voltage is usually referred to as a **voltage standing wave** or as a **stationary wave.**[1] Neither of these terms is particularly descriptive of the phenomenon. A *plot* of effective values of voltage, appearing as in Fig. 6(*e*), *is not a wave* in the usual sense. However, the term "standing wave" is in widespread use.

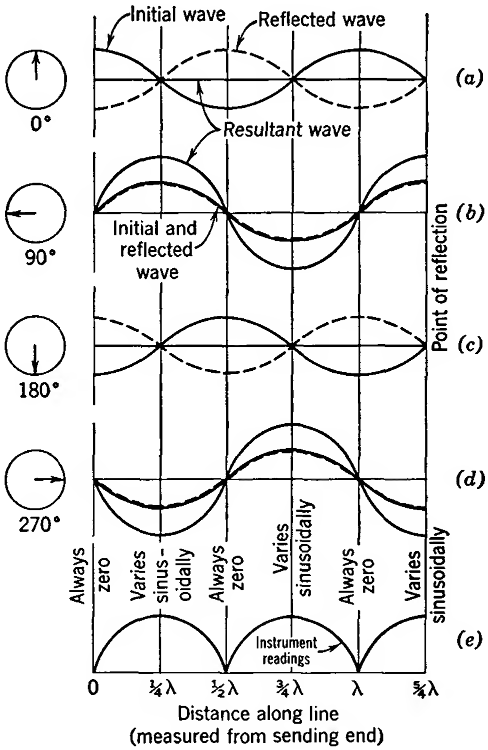


FIG. 6. Curves for electric field and voltage relations on an open-circuited lossless line $\frac{5}{4}\lambda$ in length. At $l = 0$, $l = \frac{1}{2}\lambda$, and $l = \lambda$ the resultant voltage is always zero, and at all other points the resultant voltage is as shown by the curve "instrument readings." These curves also apply to magnetic field and current distribution on a short-circuited lossless line. (Adapted from *Communication Engineering* by W. L. Everitt, McGraw-Hill Book Co.)

Voltage Distribution on a Short-Circuited Lossless Line. When an electromagnetic wave strikes the distant end of a short-circuited line, the voltage component must become zero because voltage cannot exist across zero impedance. For this to happen, it is necessary that *a 180° phase shift of the electric field and voltage component of the electromagnetic wave occur at the short-circuited end of a line.* Then, the reflected voltage component will cancel the oncoming initial voltage component, and the resultant electric field and the voltage will be zero. A short circuit provides a conducting path so that the electric charges, on which the electric lines of force terminate, can flow from one wire to the other.

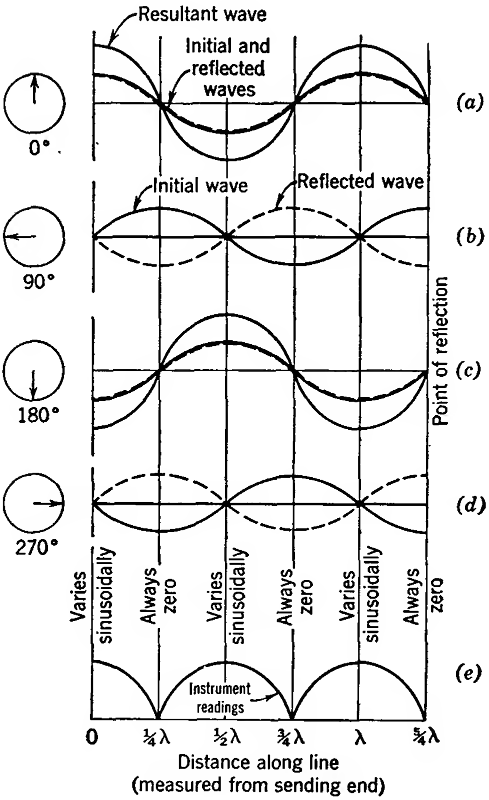


FIG. 7. Curves for electric field and voltage relations on a short-circuited lossless line $\frac{5}{4}\lambda$ in length. For (*b*) and (*d*) the resultant wave lies along the *X* axis. At certain points along the line the voltage is always zero, and at other points it is as shown by the curve "instrument readings." These curves also apply to magnetic field and current relations on open-circuited lossless lines. (See acknowledgment, Fig. 6.)

If the waves are studied at several instantaneous values of the applied generator voltage, Fig. 7 results. The various waves are combined as for Fig. 6, with the 180° angle caused by the phase reversal taken into consideration. Voltmeters connected along the line would give readings as indicated by Fig. 7(*e*), which is the "voltage standing wave," on a short-circuited lossless line. It will be noted that the voltage at the distant end always is zero, as it must be for the short-circuited line.

Current Distribution on an Open-Circuited Lossless Line. An electromagnetic wave consists of both electric field and magnetic field

components. Thus, in studying wave-reflection phenomena, current distribution, as well as voltage distribution, must be investigated, and Fig. 7 will be used.

The impressed voltage will, of course, force a current into the line. The initial current component of the electromagnetic wave will travel down the line as indicated in Fig. 7(*a*) Because the line is open circuited, the current at the distant end must be zero. For this to be possible it is necessary that *a 180° phase shift of the magnetic field and current component of the electromagnetic wave occur at the open-circuited end of the line.* Then, the reflected component will cancel the oncoming initial component, and the resultant magnetic field and the current will be zero.

If the relations between the initial and the reflected components are studied for other input phase positions, Figs. 7(*b*), (*c*), and (*d*) result. If thermomilliammeters are inserted at various points along the line, the effective value of the current readings will plot as shown in Fig. 7(*e*). This curve is called a **current standing wave.**

Current Distribution on a Short-Circuited Lossless Line. If the distant end of the line is short circuited, the current in the short circuit is a maximum value. For these phenomena to occur it is necessary that *the magnetic field and current component of an electromagnetic wave be reflected with no change in phase from a short-circuited end of a line.*

For this reason, Fig. 6 applies, and, if the relations are studied for several different initial waves, Fig. 6(*e*) results, which is the "current standing wave."

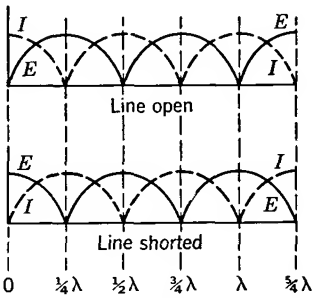


FIG. 8. Current and voltage relations on open- and short-circuited lossless lines. These are *not* initial and reflected waves. They are the (*e*) curves of Figs. 6 and 7. They are plots of the effective values of voltages and currents at various points along the line. Sending end at left, and receiving end at right.

Current and Voltage Distributions on Lossless Lines. A summary of the current and voltage distributions for a line $\frac{5}{4}\lambda$ long is made in Fig. 8. For the *lossless* line the maximum values or **antinodes** are twice the magnitude that would exist if there were no reflection, and the minimum values or **nodes** reach zero. If the line is some length other than $\frac{5}{4}\lambda$, the pattern of the voltage or current distribution can be found from Fig. 9.

Input Impedance of Open and Shorted Lossless Lines. Such lines cannot absorb power, and hence their input impedances must be

either pure inductive reactance, pure capacitive reactance, or they must be zero or infinity. These statements can be proved by applying equations 61 and 65 to lossless lines of various lengths and can be demonstrated vectorially as in Fig. 10.

Open-Circuited Line Less than $\frac{1}{4}\lambda$ in Length. It is assumed that the line is lossless and that the characteristic impedance is pure resistance. The voltage E_0 and the current I_0 at the open-circuited receiving end

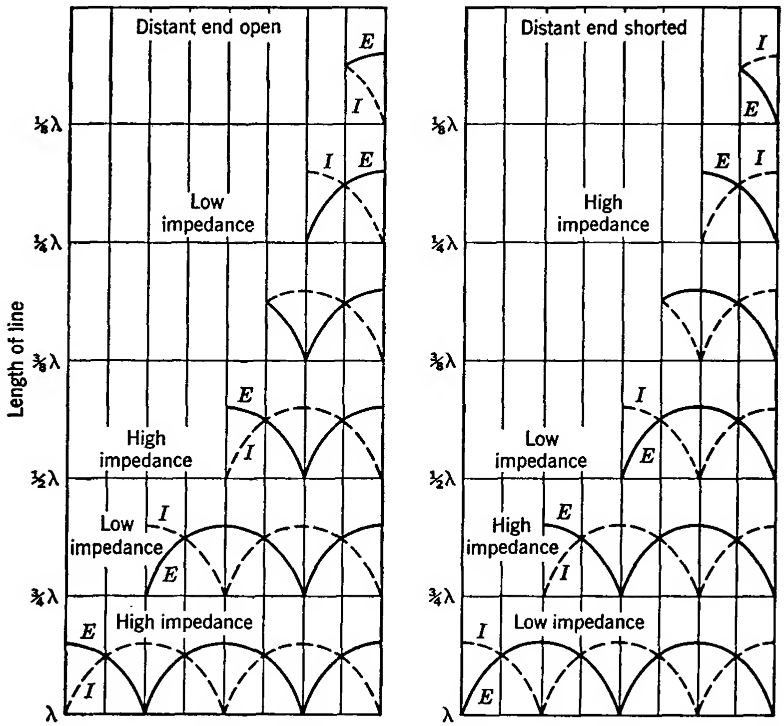


FIG. 9. Current and voltage distributions and input impedances, for open-circuited and short-circuited lossless lines of several lengths. Sending end at left, and receiving end at right.

are in phase. These have been taken as a convenient base in Fig. 10. The initial voltage wave E_{is} and the initial current wave I_{is} at the sending end will be displaced ϕ degress (less than 90° for the line under consideration) as shown. At the distant open end the current component is reflected with a change in phase of 180° and then travels toward the sending end, becoming I_{rs}. The angles ϕ and ϕ' are equal because as much time is required to travel in one direction as the other. The voltage component is reflected without a change in phase, and the

reflected component E_{rs} at the sending end has the position shown. Of course ϕ'' and ϕ are equal.

The voltage at the sending end is composed of two voltages, the initial component E_{is} is just entering the line, and the reflected component E_{rs} is just arriving back from the distant open end. The actual sending-end voltage will accordingly be the vector sum of E_{is} and E_{rs} and becomes E_s of Fig. 10. Two current components exist at the sending end, the initial component I_{is} is just entering the line, and I_{rs} is just arriving back from the distant open end. The resultant current I_s at the sending end is the vector sum of these two components.

Thus, the sending-end voltage E_s and the sending-end current I_s for a lossless line less than $\frac{1}{4}\lambda$ in length are 90° out of phase, with the current leading the voltage. The input impedance is, therefore, *a value of pure capacitive reactance, and the line is equivalent to a capacitor.* The numerical value of the reactance or the capacitance can be determined by equation 58 or by equation 61 discussed on page 206.

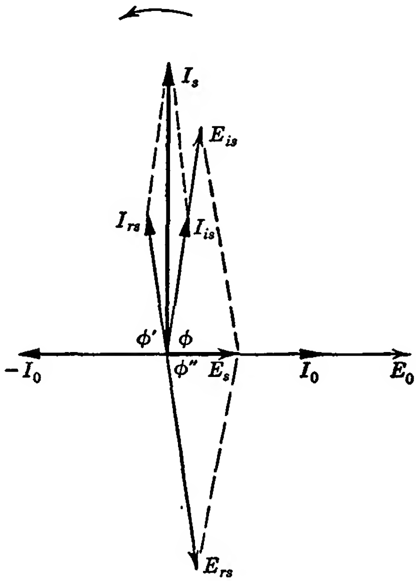


FIG. 10. Vector diagram for an open-circuited lossless line less than $\frac{1}{4}\lambda$ in length. The input impedance is pure capacitive reactance because I_s leads E_s by 90°.

Open-Circuited Line Greater than $\frac{1}{4}\lambda$ in Length. If an open-circuited lossless line has a length greater than $\frac{1}{4}\lambda$, but less than $\frac{1}{2}\lambda$, Fig. 10 can be extended to cover the conditions. Vectors E_{is} and I_{is} will lead vectors E_0 and I_0, by an angle ϕ *greater than* 90° because the line length is greater than $\frac{1}{4}\lambda$. Also, vector I_{rs} lags vector $-I_0$ by an angle ϕ' that is *greater than* 90°. Furthermore, angle ϕ'' exceeds 90°. If these changes are made on Fig. 10, E_{is} will be in the second quadrant, E_{rs} will be in the third quadrant, and the resultant voltage E_s will be opposite in direction to that shown in Fig. 10. Vector I_s will be in the same position, and hence the sending-end voltage E_s will lead the sending-end current I_s by 90°. The input impedance is, therefore, *a value of pure inductive reactance, and the line is equivalent to an*

inductor. The numerical value of the reactance or the inductance can be determined by equation 58 or by equation 61 on page 206.

Short-Circuited Line Less than $\frac{1}{4}\lambda$ in Length. A vector diagram for this line is shown in Fig. 11. Vectors E_0, I_0, E_{is}, and I_{is} are located as previously explained. But, since the line is short circuited, the voltage component is reversed at reflection as shown by $-E_0$. The reflected component at the sending end is E_{rs}, and lags $-E_0$ by the angle ϕ', and the sending-end voltage is the resultant, E_s. The reflected current component at the sending end is I_{rs} and lags I_0 by the angle ϕ''. The sending-end current is the resultant I_s, which lags the sending-end voltage E_s by 90°. The input impedance is, therefore, *a value of pure inductive reactance, and the line is equivalent to an inductor.* The numerical value of the reactance or the inductance can be determined by equation 58 or by equation 65 on page 207.

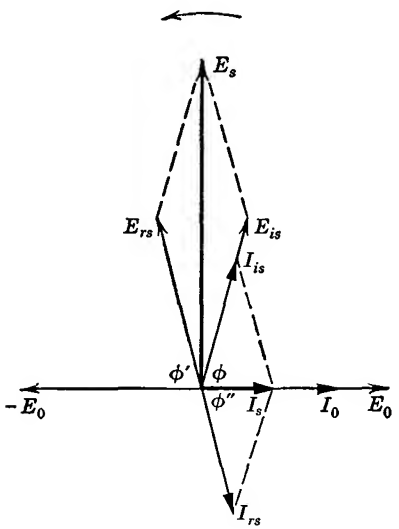


FIG. 11. Vector diagram for a short-circuited lossless line less than $\frac{1}{4}\lambda$ in length. The input impedance is pure inductive reactance because I_s lags E_s by 90°.

Short-Circuited Line Greater than $\frac{1}{4}\lambda$ in Length. If a short-circuited lossless line has a length greater than $\frac{1}{4}\lambda$, but less than $\frac{1}{2}\lambda$, Fig. 11 can be extended to give the input-impedance relations. The angles are greater than 90°, and because of this the resultant sending-end current vector will be opposite in direction from that of Fig. 11 and the resultant sending-end voltage vector will remain in the position shown. Thus, the current leads the voltage by 90°, the input impedance is *a value of pure capacitive reactance, and the line is equivalent to a capacitor.* The numerical value of the reactance or the capacitance can be determined by equation 58 or by equation 65 on page 207.

The relationships discussed in this section are summarized in Fig. 12. In studying these diagrams, line lengths should be measured from the distant end. For instance, an open-circuited line $\frac{3}{8}\lambda$ in length would have a value of reactance that was "low," and the angle would be lagging, indicating inductive reactance.

Open-circuited and short-circuited lossless lines often are called **resonant lines** because their input impedances vary between zero and

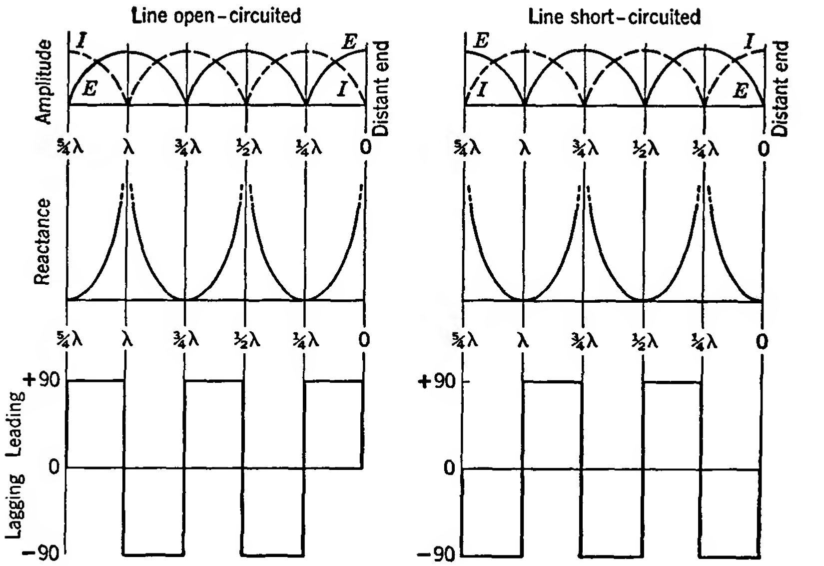


FIG. 12. Input-impedance characteristics of open-circuited and short-circuited lossless lines of various lengths. For instance, for an open-circuited line slightly less than $\frac{1}{4}\lambda$ in length, the input impedance is a low value of capacitive reactance (leading), but, if the line is slightly greater than $\frac{1}{4}\lambda$, the input impedance is a low value of inductive reactance. For the short-circuited line slightly less than $\frac{1}{4}\lambda$, the input impedance is a large value of inductive reactance (lagging), but, if the line is slightly greater than $\frac{1}{4}\lambda$, the input impedance is a large value of capacitive reactance.

infinity (Fig. 12). Such lines are extensively used to simulate (Fig. 13) capacitors, inductors, series resonant circuits (page 63), and parallel resonant circuits (page 64).

Constants of Open-Wire Lines at Low Frequencies.[9] The linear electrical constants of a transmission line were discussed on page 193. In this section the numerical values of these will be considered for frequencies particularly from about zero to 50,000 cycles. Open-wire telephone lines are usually made of hard-drawn copper wire of three standard sizes.

Series Resistance R. Series resistance R is the effective series resistance and is composed of the direct-current resistance and the additional resistance effect caused by skin effect (page 52). Theoretically, a **proximity effect** also exists, but in open-wire lines this has negligible effect on the resistance. This proximity effect is caused by the presence of the other closely parallel wires. The alternating, or effective, resistance of standard telephone lines is shown in Fig. 14.

The increase in resistance caused by skin effect is usually determined

from curves[10] such as Fig. 15. The effective resistance R_{ac} is related to the direct resistance R_{dc} by the equation

$$R_{ac} = \sigma R_{dc}, \tag{66}$$

where σ has the value given by Fig. 15. For this figure, the value of x *for copper wires* is given by the relation

$$x = 0.271\, d\sqrt{f}, \tag{67}$$

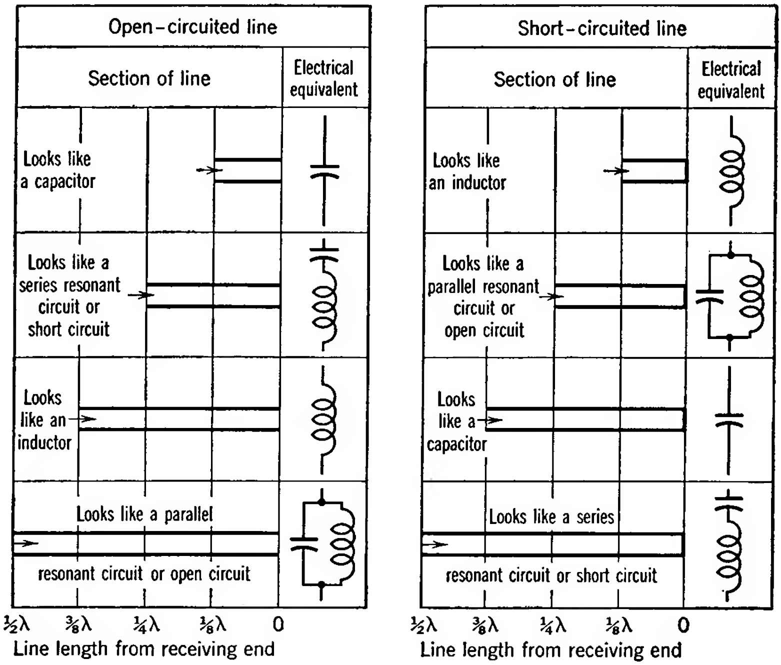


FIG. 13. Summary of input-impedance characteristics of lossless open-circuited and short-circuited lines of various lengths. The equivalent electric circuits are also shown. The input impedance of an open-circuited line less than $\lambda/4$ in length is capacitive reactance. If the length is exactly $\lambda/4$, the input impedance is zero and the line hence has the characteristics of a series resonant circuit, etc.

where d is the wire diameter in mils, and f is the frequency in megacycles.

Both the contact resistance in wire joints and the actual decrease in the size of the wire due to corrosion tend to make the measured resistance of old wires greater than that given by Fig. 14.

For the hard-drawn copper used in communication circuits, the effective resistance of the wire varies with temperature according to

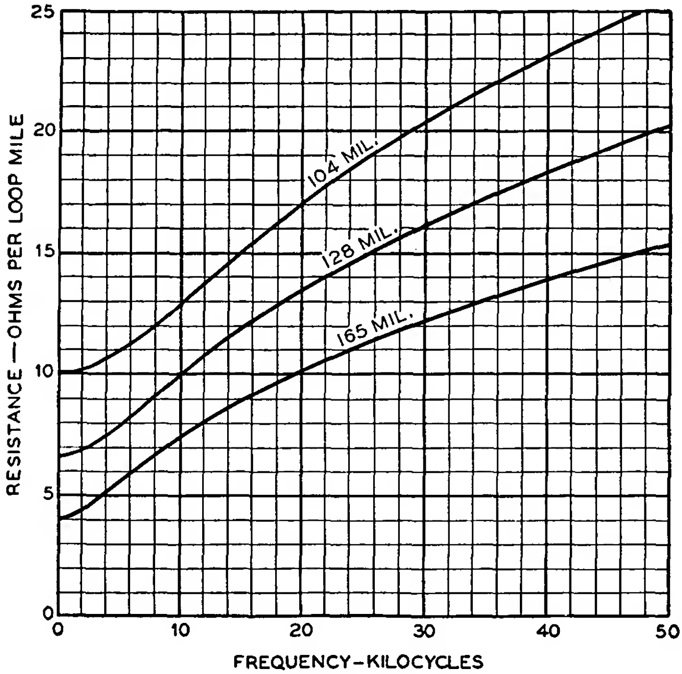


FIG. 14. Alternating-current resistance of open-wire copper pairs at 20°C (68°F). (Reference 9.)

the relation[9]

$$R = R_1[1 + A_1(t - t_1)], \tag{68}$$

where R is the effective resistance at temperature t, R_1 is the resistance at temperature t_1, and A_1 is the alternating-current temperature coefficient of resistance at temperature t_1 and can be obtained from Fig. 16. Temperatures are in degrees centigrade.

Series Inductance L. The series self-inductance of an open-wire transmission line of two parallel wires is[9]

$$L = 0.64374\left(2.3026 \log_{10} \frac{2D}{d} + \mu\delta\right) \times 10^{-3}, \tag{69}$$

where L is in henrys per loop mile, d is the diameter of each wire, and D is the distance between centers both in the same units, μ is the permeability of the wire, and δ is a factor depending on frequency.[11] For hard-drawn copper wires, the permeability is unity, and for audio frequencies[9] $\mu\delta = 0.25$. Equation 69 applies with best accuracy when

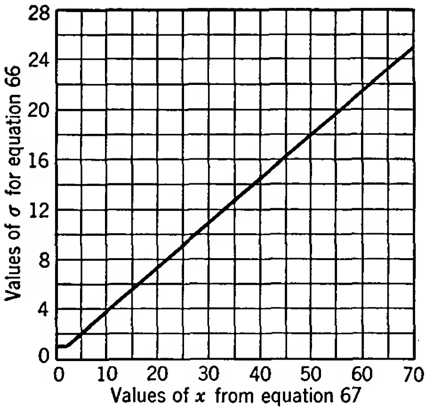


FIG. 15. Values of σ for equation 66 to obtain the alternating-current resistance of copper wires, particularly at radio frequencies. For hard-drawn copper used in open-wire lines the resistance is about 2 or 3 per cent greater than for annealed copper.

the ratio D/d is at least 10. The inductance of telephone lines at various frequencies is given in Table I.

Shunt Capacitance C. The shunt capacitance of an open-wire transmission line of two parallel wires is[9]

$$C = \frac{0.019415}{\log_{10} \frac{2D}{d}} \times 10^{-6}, \quad (70)$$

where C is in farads per loop mile (mile of *line*), d is the diameter, and D is the distance between centers both in the same units. This equation applies with greatest accuracy when the ratio D/d is greater than about 10.

Values for the capacitance between wires of different sizes and of different spacings are given in Table II. Since on an actual line the

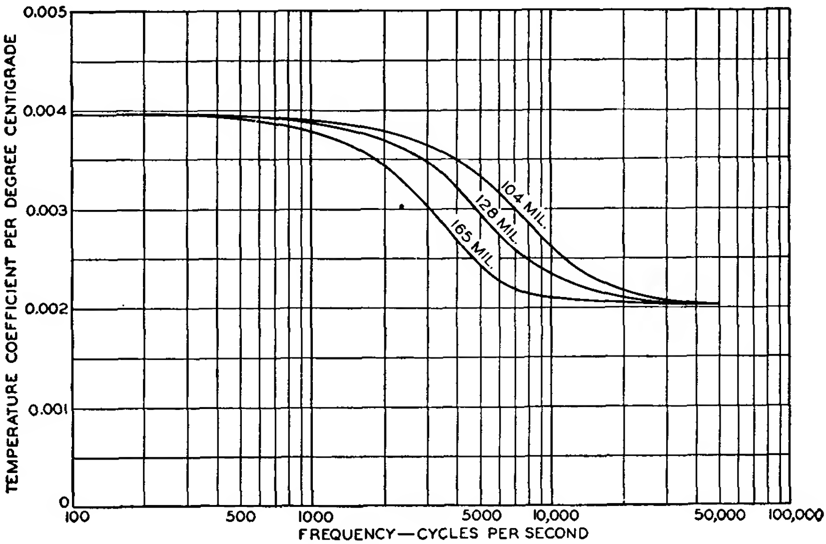


FIG. 16. Alternating-current temperature coefficient of resistance for open-wire pairs at 20°C. (Reference 9.)

TABLE I

INDUCTANCE OF OPEN-WIRE LINES

(From reference 9.)

Frequency, Cycles per Second	Inductance of Open-Wire Pairs, Henrys per Loop Mile								
	165-mil			128-mil			104-mil		
	8-in.	12-in.	18.25-in.	8-in.	12-in.	18.25-in.	8-in.	12-in.	18.25-in.
0	0.00311	0.00337	0.00364	0.00327	0.00353	0.00380	0.00340	0.00366	0.00393
1,000	0.00311	0.00337	0.00364	0.00327	0.00353	0.00380	0.00340	0.00366	0.00393
10,000	0.00305	0.00331	0.00358	0.00323	0.00349	0.00376	0.00338	0.00364	0.00391
25,000	0.00301	0.00327	0.00354	0.00319	0.00345	0.00372	0.00334	0.00360	0.00387
50,000	0.00299	0.00325	0.00352	0.00317	0.00343	0.00370	0.00331	0.00357	0.00384
Infinite	0.00295	0.00321	0.00348	0.00311	0.00337	0.00364	0.00324	0.00350	0.00377

capacitance depends to some extent on the presence of other wires,[9] values are given for wires isolated in space, and also for wires on a 40-wire line. The effect of frequency is negligible.

TABLE II

CAPACITANCE VALUES FOR OPEN-WIRE LINES

(From reference 9.)

Wire Spacing	Capacitance—Microfarads per Mile					
	165-mil		128-mil		104-mil	
	In space	On 40-wire line	In space	On 40-wire line	In space	On 40-wire line
8 in.	0.00977	0.00996	0.00926	0.00944	0.00888	0.00905
12 in.	0.00898	0.00915	0.00855	0.00871	0.00822	0.00837
18.25 in.	0.00828	0.00863	0.00791	0.00825	0.00763	0.00797

Shunt Conductance G. This is the most erratic of the line constants.[9] Shunt conductance is *not* the direct-current conductance between wires but is of the nature explained[9] in the following quotation:

The determination of the value of G for direct current is quite simple, involving merely a measurement of the actual conductance between wires for a length of circuit short enough to avoid propagation effects. For alternating currents, however, it is customary to employ an equivalent value of G which

includes all the losses suffered by the power transmitted over the pair except the normal I^2R loss in the wires themselves. This inclusion of numerous little-understood losses in the general term leakage has at times served to insulate the individual losses from analysis.*

There are many paths by which the leakage currents can reach the other wire of a pair, or ground. Leakage occurs through the air itself but is of little importance. It occurs through trees or other foreign objects touching the wires. It also occurs at insulators. Insulator leakages are discussed in a paper[12] from which the following is summarized.

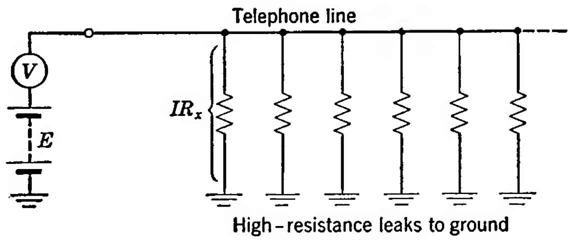


FIG. 17. Illustrating the voltmeter method of measuring the direct-current leakage, or insulation, resistance of a line to ground.

Current leakage at insulators may be resolved into two components, one in phase with the potential and one leading the potential by 90°. This in-phase component represents an energy loss and is the more important component. The term *shunt conductance* refers to the conductance corresponding to the in-phase component.

Both the in-phase and reactive components in flowing through the resistance of the line conductors produce energy losses, but these are negligible. All other energy losses that occur because of the insulators, whether they actually occur at the insulators or elsewhere, are attributed to the insulators.

Shunt conductance varies greatly with weather conditions, the age of the insulators and the line, and similar factors. It is extremely difficult to maintain proper insulation on lines close to the sea or to salt lakes,[13] or in certain regions subject to alkali dust storms. Methods of measuring the shunt conductance G are discussed in reference 14. Values are given in Table IV.

Lines are quite often given direct-current tests to determine their direct-current insulation resistance. Such tests are often made with a voltmeter as indicated in Fig. 17. A voltmeter having a resistance of

* Reprinted by permission, courtesy *Bell System Technical Journal.*

about 100,000 ohms is satisfactory. The IR_x drop across the insulation resistance of the entire line equals the battery voltage E minus the reading V of the voltmeter, or $IR_x = E - V$, and $R_x = (E - V)/I$. The voltmeter and the equivalent resistance of the line are in series, and thus the current through each is the same. The current through the voltmeter (and R_x) is $I = V/R_v$, and thus it can be written that

$$R_x = \frac{(E - V)R_v}{V}. \tag{71}$$

Open-Wire Telephone Lines at Audio Frequencies. Data applying to telephone lines are summarized in Tables III and IV.

TABLE III

CHARACTERISTICS OF EXCHANGE OPEN-WIRE LINES

Type of Circuit	Diameter of Wires in Inches	Size of Wires	Resistance R Ohms per Loop Mile, 1000 Cycles, 68° F	Inductance L Henrys per Loop Mile, 1000 Cycles	Capacitance C Microfarads per Loop Mile	Characteristic Impedance Magnitude in Ohms, 1000 Cycles	Attenuation Decibels per Mile at 1000 Cycles (Dry Weather)
Copper	0.080	No. 14NBS	17.1	0.00373	0.00783	768	0.103
Copper	0.104	No. 12NBS	10.2	0.00365	0.00837	691	0.066
Iron	0.083	No. 14BWG	134.4	0.01626	0.00793	1841	0.356
Iron	0.109	No. 12BWG	93.8	0.01469	0.00847	1573	0.283

The use of Table IV and the equations of the preceding section will now be considered. A non-pole pair side circuit of hard-drawn 165-mil copper wires spaced 12 inches apart will be used as an illustration, the frequency will be 1000 cycles, and the temperature 20°C.

Calculation of Linear Electrical Constants. The *series resistance R* is found as follows: $R_{\text{dc}} = \rho l/d^2 = 10.37 \times 2 \times 5280/165^2 = 4.02$ ohms per loop mile. For hard-drawn copper the resistance is assumed 3 per cent greater; hence, $R_{\text{dc}} = 4.02 \times 1.03 = 4.14$ ohms per loop mile. From equation 67, $x = 0.271\, d\sqrt{f} = 0.271 \times 165 \times \sqrt{1000 \times 10^{-6}} = 1.41$, and, when this is applied to Fig. 15, it is seen that the skin effect is negligible at 1000 cycles. Table IV gives 4.11 ohms.

The *series inductance L* is calculated by equation 69, $L = 0.64374 \left(2.3026 \log_{10} \frac{2D}{d} + \mu\delta\right) \times 10^{-3}$. Evaluating the portion $\log_{10} \frac{2D}{d}$ gives $\log_{10} \frac{2 \times 12}{0.165} = 2.163$. As stated on page 217, $\mu\delta = 0.25$ for copper at audio frequencies. Thus, the inductance is $L = 0.64374(2.3026 \times$

TABLE IV

Characteristics of Standard Types of Open-Wire Long-Distance Toll Telephone Circuits
Copper Wire—1000 Cycles per Second

Type of Circuit	Gauge of Wire, Mils	Spacing of Wires, Inches	Constants per Loop-Mile				Propagation Constant per Mile				Characteristic Impedance				Wave-length, Miles	Velocity, Miles per Second	Attenuation, Decibel per Mile
							Polar		Rectangular		Polar		Rectangular				
			R Ohms	L Henrys	C Micro-farads	G Micro-mhos	Magnitude	Angle Degrees (Pos.)	α Nepers	β Radians	Magnitude Ohms	Angle Degrees (Neg.)	R Ohms	X Ohms (Neg.)			
Non-Pole Pair Side	165	12	4.11	0.00337	0.00915	0.29	0.0352	84.36	0.00346	0.0350	612	5.35	610	57	179.5	179,500	0.0300
Pole Pair Side	165	18	4.11	0.00364	0.00863	0.29	0.0355	84.75	0.00325	0.0353	653	5.00	651	57	178.0	178,000	0.0282
Non-Pole Pair Phantom	165	12	2.06	0.00208	0.01514	0.58	0.0355	85.34	0.00288	0.0354	373	4.30	372	28	177.5	177,500	0.0250
Pole Pair Phantom	165	18	2.06	0.00207	0.01563	0.58	0.0359	85.33	0.00293	0.0358	366	4.33	365	28	175.5	177,500	0.0254
Non-Pole Pair Physical	165	8	4.11	0.00311	0.00996	0.14	0.0353	83.99	0.00370	0.0351	565	5.88	562	58	179.0	179,000	0.0321
Non-Pole Pair Side	128	12	6.74	0.00353	0.00871	0.29	0.0356	81.39	0.00533	0.0352	650	8.32	643	94	178.5	178,500	0.0462
Pole Pair Side	128	18	6.74	0.00380	0.00825	0.29	0.0358	81.95	0.00502	0.0355	693	7.72	686	93	177.0	177,000	0.0436
Non-Pole Pair Phantom	128	12	3.37	0.00216	0.01454	0.58	0.0357	82.84	0.00445	0.0355	401	6.73	398	47	177.0	177,000	0.0386
Pole Pair Phantom	128	18	3.37	0.00215	0.01501	0.58	0.0362	82.82	0.00453	0.0359	384	6.83	382	46	174.8	174,800	0.0393
Non-Pole Pair Physical	128	8	6.74	0.00327	0.00944	0.14	0.0358	80.85	0.00569	0.0353	603	8.97	596	94	178.0	178,000	0.0495
Non-Pole Pair Side	104	12	10.15	0.00366	0.00837	0.29	0.0363	77.93	0.00760	0.0355	692	11.75	677	141	177.0	177,000	0.0660
Pole Pair Side	104	18	10.15	0.00393	0.00797	0.29	0.0365	78.66	0.00718	0.0358	730	10.97	717	139	175.5	175,500	0.0624
Non-Pole Pair Phantom	104	12	5.08	0.00223	0.01409	0.58	0.0363	79.84	0.00640	0.0357	421	9.70	415	71	176.0	176,000	0.0556
Pole Pair Phantom	104	18	5.08	0.00222	0.01454	0.58	0.0368	79.81	0.00651	0.0362	403	9.83	397	69	173.6	173,600	0.0565
Non-Pole Pair Physical	104	8	10.15	0.00340	0.00905	0.14	0.0367	77.22	0.00811	0.0358	644	12.63	629	141	175.5	175,500	0.0704

Notes: 1. All values are for dry weather conditions.
2. All capacitance values assume a line carrying 40 wires.
3. Resistance values are for temperature of 20° C (68° F).
4. DP Insulators assumed for all 12-inch and 18-inch spaced wires—CS Insulators for all 8-inch spaced wire.
5. Open-wire lines are no longer loaded in the United States.

$2.163 + 0.25) \times 10^{-3} = 3.37 \times 10^{-3}$ henry, or 0.00337 henry per loop mile, or per mile of line. This agrees with Table IV.

The *shunt capacitance* C is calculated by equation 70, $C = [0.019415/\log_{10}(2D/d)] \times 10^{-6}$, which for the line under consideration is $C = [0.019415/\log_{10}(24/0.165)] \times 10^{-6} = 0.00898 \times 10^{-6}$ farad, or 0.00898 microfarad per mile. This is the value of the capacitance in free space and will not agree exactly with Table IV, which is for pairs on a 40-wire line.

The *shunt conductance* G is not calculated. From Table IV, the value is 0.29 micromho per mile.

Calculation of Propagation Constant. The actual constants from Table IV, instead of the constants just calculated, will be used for subsequent calculations. These are $R = 4.11$ ohms, $L = 0.00337$ henry, $C = 0.00915 \times 10^{-6}$ farad, and $G = 0.29 \times 10^{-6}$ mho. At 1000 cycles, and from equation 37,

$$
\begin{aligned}
\gamma = \alpha + j\beta &= \sqrt{(R+jX)(G+jB)} \\
&= \sqrt{(4.11+j6.28\times1000\times0.00337)(0.29\times10^{-6}+j6.28\times1000\times0.00915\times10^{-6})} \\
&= \sqrt{(4.11+j21.2)(0.29\times10^{-6}+j57.5\times10^{-6})} \\
&= \sqrt{(21.6\underline{/79^\circ})(57.5\times10^{-6}\underline{/90^\circ})} \\
&= 0.0353\underline{/84.5^\circ} = 0.00339+j0.0352.
\end{aligned}
$$

Thus, the attenuation constant $\alpha = 0.00339$ neper, or $0.00339 \times 8.686 = 0.0295$ decibel per mile. The phase constant $\beta = 0.0352$ radian or $0.0352 \times 57.3^\circ = 2.01^\circ$ per mile. These values agree approximately with Table IV.

Calculation of Characteristic Impedance. From equation 50,

$$
\begin{aligned}
Z_O &= \sqrt{(R+jX)/(G+jB)} = \sqrt{(21.6\underline{/79^\circ})/(57.5 \times 10^{-6}\underline{/90^\circ})} \\
&= 610\underline{/-5.5^\circ} = 607 - j58 \text{ ohms.}
\end{aligned}
$$

Calculation of Wavelength and Wave Velocity. From equation 54, $\lambda = 2\pi/\beta = 6.28/0.0352 = 179$ miles; or, $\lambda = 360^\circ/2.01^\circ = 179$ miles. From equation 55, $V = \omega/\beta = 6.28 \times 1000/0.0352 = 179{,}000$ miles per second, or $V = \lambda f = 179 \times 1000 = 179{,}000$ miles per second.

Calculation of Line Performance. A power input of 0.001 watt, or 1.0 milliwatt, at 1000 cycles is the standard testing power used in checking the performance of telephone lines. Calculations will be made on a section of line 250 miles long, and terminated in 610 ohms pure resistance, a value that simulates the characteristic impedance Z_O sufficiently close for practical purposes. It will be assumed that the line input impedance also is 610 ohms resistance.

The *input voltage* will be $E = \sqrt{PR} = \sqrt{0.001 \times 610} = 0.782$ volt.

The *input current* will be $I = \sqrt{P/R} = \sqrt{0.001/610} = 0.00128$ ampere, or $I = E/R = 0.782/610 = 0.00128$ ampere.

The *received power* can be found from equation 49, page 86, $P_r = P_s 10^{-0.1ln} = 1.0 \times 10^{-0.1\times 7.5} = 0.178$ milliwatt, or 0.000178 watt. The value $ln = 7.5$ is the loss in decibels for the 250-mile section of line having a loss of 0.03 decibel per mile from Table IV.

The *received voltage* can be found from equation 53, $E_r = E_s 10^{-0.05ln} = 0.782 \times 10^{-0.05\times 7.5} = 0.33$ volt. Or the received voltage is $\sqrt{0.000178} \times 610 = 0.33$ volt.

The *received current* can be found from equation 52, $I_r = I_s 10^{-0.05ln} = 0.00128 \times 10^{-0.05\times 7.5} = 0.00054$ ampere. Or the received current is $0.33/610 = 0.00054$ ampere or 0.54 milliampere.

The loss on a typical open-wire line at frequencies higher than 1000 cycles is given in Fig. 18.

The Phantom Circuit. Reference is made in Table IV to the phantom circuit. This circuit is obtained, as indicated in Fig. 19, from two pairs of line wires (or from two cable pairs). By the use of the phantom circuit, *three* telephone conversations can be carried on simultaneously over two pairs of wires.

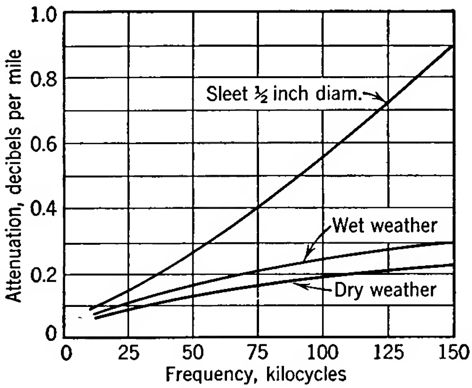


FIG. 18. Typical losses on a 165-mil open-wire copper telephone line at various frequencies. (From *Bell System Tech. J.*, January, 1939, Vol. 18, No. 1.)

If the transformers (or **repeating coils,** as they are commonly called in telephony) are tapped at exactly their electrical center, and if the impedance of each line wire is the same, the currents from the phantom circuit will divide equally. Since these currents will flow in opposite directions through the transformer windings, their magnetic effects will neutralize, and no magnetic flux will be produced to induce speech currents in the side circuits. Instantaneous directions of current flow for a phantom circuit are shown; side-circuit currents will flow as if the phantom circuit were not present.

Constants of Open-Wire Lines at Radio Frequencies.[15, 16, 17] Such lines are used in radio for connecting the radio transmitter to

the sending antenna and for similar purposes. It is assumed that the conductors are of hard-drawn copper.

Series Resistance R. The resistance at radio frequencies of *both* wires of a transmission line is

$$R = \frac{16.8\sqrt{f}}{d}, \tag{72}$$

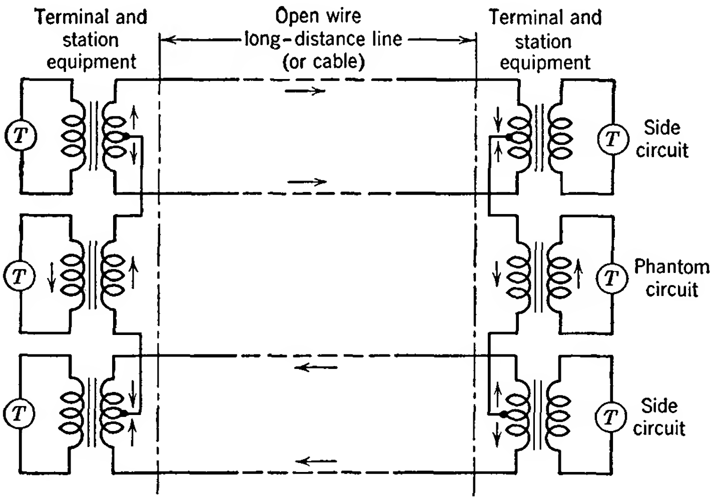


FIG. 19. A phantom telephone circuit is obtained over four wires as shown. The letter T signifies a telephone set.

where f is the frequency in cycles per second, d is the diameter of the wires in centimeters, and R is in microhms per meter of line. The resistance can also be found by the method of page 215.

Series Inductance L. The series self-inductance at radio frequencies of *both* wires of a transmission line is

$$L = 0.921 \log_{10} \frac{2D}{d}, \tag{73}$$

where D is the distance between the wires, d is the diameter of the wires in the same units as D, and L is in microhenrys per meter of line.

Shunt Capacitance C. The shunt capacitance between the two wires of a radio-frequency transmission line is

$$C = \frac{0.000012}{\log_{10} \frac{2D}{d}}, \tag{74}$$

where C is the capacitance in microfarads per meter of transmission line and D and d are as previously explained. Equations 73 and 74 do not apply for twisted pairs and are accurate only when D/d is about 10 or more.

Shunt Conductance G. Shunt conductance G was discussed on page 219 for an open-wire line at low frequencies. A similar discussion applies at high frequencies, but this constant is usually neglected in radio.

Propagation Constant of Radio-Frequency Transmission Lines. The propagation constant of any transmission line at *any* frequency is given by equation 37. The attenuation constant α and the phase constant β can be found by equations 43 and 49. At radio frequencies, however, ωL and ωC become so much greater than the resistance R and conductance G that simplifications are possible. Thus, equation 37 becomes[5, 15]

$$\gamma = \alpha + j\beta = \left(\frac{R}{2}\sqrt{\frac{C}{L}} + \frac{G}{2}\sqrt{\frac{L}{C}}\right) + j\omega\sqrt{LC}, \tag{75}$$

and, when the shunt conductance is negligible,

$$\alpha = \frac{R}{2}\sqrt{\frac{C}{L}} = \frac{R}{2Z_O}, \tag{76}$$

and

$$\beta = \omega\sqrt{LC}, \tag{77}$$

where α is the attenuation constant in nepers per unit length, β is the phase constant in radians per unit length (corresponding to R, L, and C), Z_O is the characteristic impedance (equation 79), and the units ohms, farads, and henrys are used. Since the wave velocity is, from equation 55, $V = \omega/\beta$, it follows that at radio frequencies the velocity of propagation on a line is approximately

$$V = \frac{\omega}{\beta} = \frac{\omega}{\omega\sqrt{LC}} = \frac{1}{\sqrt{LC}}. \tag{78}$$

This is the velocity of light, approximately 186,300 miles per second, or 3×10^5 kilometers, 3×10^8 meters, or 3×10^{10} centimeters per second.

Characteristic Impedance of Radio-Frequency Transmission Lines. The characteristic impedance of any line at any frequency is given by equation 50. At radio frequencies ωL and ωC become so much

greater than R and G that they may be neglected, and accordingly

$$Z_O = \sqrt{\frac{R + j\omega L}{G + j\omega C}} = \sqrt{\frac{L}{C}}, \tag{79}$$

where L is the series inductance of the line in henrys, C is the shunt capacitance; Z_O will be the characteristic impedance in ohms and will be *pure resistance*. The values of L and C may be for any (but the same) length of line. If equations 73 and 74 are substituted in equation 79, the characteristic impedance of a radio-frequency line becomes

$$Z_O = 276 \log_{10} \frac{2D}{d}, \tag{80}$$

where Z_O is ohms resistance.

A helpful set of radio-frequency transmission-line nomographs, prepared by P. H. Smith, has been published in the February, 1949, issue of *Electronics*, Vol. 22, No. 2.

Open-Wire Transmission Line at Radio Frequencies. The methods of using preceding equations will be considered now. As an illustration a line composed of two No. 4 A.W.G. (diameter, 0.518 centimeter) hard-drawn copper wires spaced 18 inches (45.7 centimeters) apart will be considered at a frequency of 10.0 megacycles and at 20°C.

Calculation of Linear Electrical Constants. The *series resistance* R is, from equation 72, $R = 16.8\sqrt{10^7}/0.518 = 102.5 \times 10^3$ microhms per meter, or 0.1025 ohm per meter.

The *series inductance* L is, from equation 73, $L = 0.921 \log_{10} 2 \times 45.7/0.518 = 2.065$ microhenrys per meter.

The *shunt capacitance* C is, from equation 74, $C = 0.000012/\log_{10} 2 \times 45.7/0.518 = 0.00000534$ microfarad per meter.

The *shunt conductance* G is assumed negligible.

Calculation of Propagation Constant. Propagation constant may be calculated by equations 37 or 75, or α and β may be evaluated by equations 76 and 77. Using these last two equations, $\alpha = (0.1025/2) \times \sqrt{0.00000534 \times 10^{-6}/2.065 \times 10^{-6}} = 8.26 \times 10^{-5}$ neper per meter, and $\beta = 6.28 \times 10^7\sqrt{2.065 \times 10^{-6} \times 0.00000534 \times 10^{-6}} = 0.209$ radian per meter. Using the value of Z_O calculated in the next paragraph and the second expression of equation 76, $\alpha = 0.1025/2 \times 620 = 8.27 \times 10^{-5}$ neper per meter.

Calculation of Characteristic Impedance. Characteristic impedance may be calculated from equations 50, 79, or 80. Using equation 79, $Z_O = \sqrt{2.065 \times 10^{-6}/0.00000534 \times 10^{-6}} = 620$ ohms resistance. Using

equation 80, $Z_O = 276 \log_{10} 2 \times 45.7/0.518 = 620$ ohms resistance. The characteristic impedance of typical radio-frequency transmission lines is given in Fig. 20.

Calculation of Wavelength and Wave Velocity. From equation 54, $\lambda = 6.28/\beta = 6.28/0.209 = 30$ meters (approximately). The wave velocity is, from equation 55, $V = \omega/\beta = 6.28 \times 10^7/0.209 = 3 \times 10^8$ meters per second.

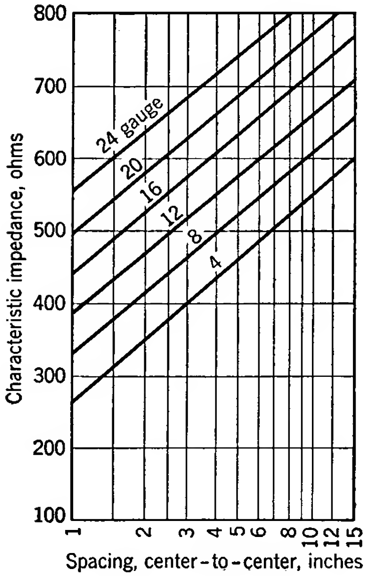


FIG. 20. Characteristic impedances in ohms (resistance) for open-wire transmission lines at radio frequencies.

Calculation of Line Performance. The radio-frequency transmission line of the preceding paragraphs is to deliver 10.0 kilowatts of power to an antenna 500 feet (152.5 meters) from the transmitter. The antenna terminates the line so that no standing waves exist. The sending-end conditions are desired.

The *receiving-end current* will be $I_r = \sqrt{P/R} = \sqrt{10{,}000/620} = 4.02$ amperes.

The *receiving-end voltage* will be $E_r = \sqrt{PR} = \sqrt{10{,}000 \times 620} = 2490$ volts.

The *sending-end power* will be computed as follows: The loss in the 500-foot (152.5-meter) line was computed to be 8.26×10^{-5} neper per meter, and hence the total loss will be about 12.6×10^{-3} neper. This is $12.6 \times 10^{-3} \times 8.686 = 0.1093$ decibel. From equation 49, page 86, the power input will be $P = 10 \times 10^{0.1 \times 0.1093} = 10.255$ kilowatts, or 10,255 watts.

The *sending-end current* will be $I_s = \sqrt{10{,}255/620} = 4.07$ amperes.

The *sending-end voltage* will be $E_s = \sqrt{10{,}255 \times 620} = 2525$ volts.

The *efficiency* of the radio-frequency transmission line will be $10{,}000/10{,}255 = 97.5$ per cent.

The preceding calculations neglect the power radiated by the radio-frequency transmission line. These **radiation losses** can be calculated approximately by the equation[15]

$$P = 160\ (I\pi D/\lambda)^2, \tag{81}$$

where P is the radiation loss in watts, I is the line current in amperes, and D is the line spacing and λ the wavelength, both in the same units. This equation holds[15] if the line is terminated in its characteristic impedance, if the line spacing is less than 0.1λ and if the line length is more than $20D$.

The Exponential Radio-Frequency Transmission Line. The spacing of the wires of the transmission lines previously considered has been the same at all points, and hence the inductance and capacitance of each unit length is the same. An **exponential transmission line**[16, 17, 18] is one in which the inductance and the capacitance vary exponentially with distance along the line. Such a variation in these line constants can be achieved by varying the spacing exponentially, by varying the wire size, or by these two methods in combination.

It has been found[18, 19, 20] that the exponential line has impedance-matching properties. In this respect the exponential line is similar to the non-symmetrical network which must be terminated in *image impedances* that are unequal (page 153). Also, the exponential line has a cutoff frequency. In these respects it differs from the uniform line. The exponential line is used at radio frequencies for impedance-matching purposes (page 495). See also page 498.

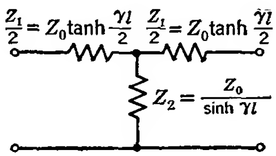


FIG. 21. Equivalent T network for a uniform transmission line or non-loaded cable.

Equivalent T Section of a Uniform Line. It was shown on page 146 that any complex network could be represented by a T section (or a π section). Thus, using equations 19, 20, and 21 of Chapter 5, the equivalent T section for a transmission line can be determined. The equivalent T section (Fig. 21) for a transmission line having uniformly distributed constants can be calculated directly if the characteristic impedance Z_O and the propagation constant per section γ are known.

On page 206 the impedance measured at the sending terminals of an open-circuited line was shown to be $Z_{oc} = Z_O \coth \gamma l$, and on page 207 the impedance of a short-circuited line was $Z_{sc} = Z_O \tanh \gamma l$. From these relations,

$$\tanh \gamma l = \sqrt{\frac{Z_{sc}}{Z_{oc}}} = \sqrt{\frac{Z_O \tanh \gamma l}{Z_O \coth \gamma l}}, \tag{82}$$

because, from hyperbolic trigonometry, $\coth x = 1/\tanh x$.

In terms of the symmetrical T section of Fig. 21, equation 82

becomes, from equations 47 and 48, page 156,

$$\tanh \gamma l = \sqrt{\frac{Z_{sc}}{Z_{oc}}} = \frac{\sqrt{Z_1 Z_2 + \dfrac{Z_1^2}{4}}}{\sqrt{\left(\dfrac{Z_1}{2} + Z_2\right)^2}} = \frac{Z_O}{\dfrac{Z_1}{2} + Z_2}, \tag{83}$$

when simplified by substituting the value* of Z_O from equation 49, page 156 (page 340 of reference 7). From equation 83,

$$\frac{Z_1}{2} = \frac{Z_O}{\tanh \gamma l} - Z_2, \quad \text{and} \quad Z_1 = \frac{2Z_O}{\tanh \gamma l} - 2Z_2, \tag{84}$$

and this equation contains Z_O, which is given in terms of Z_1 and Z_2 by equation 49, page 156. Substituting this value for Z_1 in $Z_O = \frac{1}{2}\sqrt{4Z_1Z_2 + Z_1^2}$ obtained from equation 49,

$$\left.\begin{aligned} Z_O &= \tfrac{1}{2}\sqrt{4Z_2\left(\frac{2Z_O}{\tanh \gamma l} - 2Z_2\right) + \left(\frac{2Z_O}{\tanh \gamma l} - 2Z_2\right)^2} \\ \text{and} \quad & \\ Z_O^2 &= \frac{Z_O^2}{\tanh^2 \gamma l} - Z_2^2. \end{aligned}\right\} \tag{85}$$

Solving, and substituting $\coth x = 1/\tanh x$, $\operatorname{csch} x = \sqrt{\coth^2 x - 1}$, and $\operatorname{csch} x = 1/\sinh x$,

$$Z_2 = Z_O\sqrt{\coth^2 \gamma l - 1} = Z_O/\sinh \gamma l. \tag{86}$$

Substituting equation 86 in equation 84,

$$Z_1 = \frac{2Z_O}{\tanh \gamma l} - \frac{2Z_O}{\sinh \gamma l} = \frac{2Z_O(\cosh \gamma l - 1)}{\sinh \gamma l} = 2Z_O \tanh \frac{\gamma l}{2}, \tag{87}$$

since $\tanh x = \dfrac{\sinh x}{\cosh x}$, and $\dfrac{(\cosh x - 1)}{\sinh x} = \tanh \frac{1}{2} x$.

Equations 86 and 87 give the values of Z_1 and Z_2 for the T network of Fig. 21 equivalent at one frequency to a uniform transmission line having distributed constants. Similar equations can also be derived for a π equivalent network.

Determination of Line Constants from Impedance Measurements.[6, 8] From equation 35, $\gamma = \sqrt{zy}$, and, from equation 50, $Z_O = \sqrt{z/y}$.

* The notation Z_O (characteristic impedance) instead of Z_K (iterative impedance) is used because a uniform line is under consideration.

Hence,

$$\gamma Z_O = z = R + j\omega L, \tag{88}$$

and

$$\gamma / Z_O = y = G + j\omega C. \tag{89}$$

From equation 51, $Z_O = \sqrt{Z_{oc} Z_{sc}}$ and can readily be found from open-circuit and short-circuit impedance measurements. Also, from equation 82 $\tanh \gamma l = \sqrt{Z_{sc}/Z_{oc}}$. These equations can be used to find the line constants as follows:

For an open-wire line 197 miles long, $Z_{oc} = 672 - j214$ or $705 \underline{/-17.65°}$ ohms, and $Z_{sc} = 695 - j75.5$ or $699 \underline{/-6.2°}$ ohms. Thus, $Z_O = \sqrt{(705 \underline{/-17.65°})(699 \underline{/-6.2°})} = 702 \underline{/-11.9°}$ ohms, and these measurements are at 1000 cycles. The value of $\tanh \gamma l =$ $\sqrt{(699 \underline{/-6.2°})/(705 \underline{/-17.65°})} = 0.996 \underline{/+5.725°} = 0.991 + j0.0992.$

The next step involves finding αl and βl from the expression for $\tanh \gamma l$. From hyperbolic trigonometry, $\tanh \gamma = \tanh (\alpha + j\beta) = A + jB$,

$$\tanh 2\alpha l = \frac{2A}{1 + A^2 + B^2}, \tag{90}$$

and

$$\tan 2\beta l = \frac{2B}{1 - (A^2 + B^2)}. \tag{91}$$

Thus, $\tanh 2\alpha l = (2 \times 0.991)/(1 + 0.991^2 + 0.0992^2) = 0.995$. From tables of hyperbolic functions, $2\alpha l = 2.99$, and $\alpha l = 1.495$ nepers. Hence for the 197-mile line, $\alpha = 1.495/197 = 0.0076$ neper per mile or $0.0076 \times 8.686 = 0.0659$ decibel per mile. Similarly, $\tan 2\beta l = (2 \times 0.0992)/[1 - (0.991^2 + 0.0992^2)] = -24.32$. From tables of circular functions, $2\beta l = 807.65°$, and $\beta l = 403.83°$. Of course, this fact must be determined from a study of the frequency and the velocity of propagation. At 1000 cycles, for a line 200 miles long and a velocity of propagation of about 180,000 miles per second, βl would be somewhere in the fifth quadrant. The value in radians is $\beta l = 408.5°/57.3° = 7.04$ radians, and $\beta = 7.04/197 = 0.0357$ radian per mile. Thus, $\gamma = \alpha + j\beta = 0.0076 + j0.0357 = 0.0365 \underline{/77.98°}$.

From equation 88,

$$\begin{aligned}\gamma Z_O = z = R + j\omega L &= 0.0365 \underline{/77.98°} \times 702 \underline{/-11.9°} \\ &= 25.6 \underline{/66.08°} = 10.38 + j23.41 \text{ ohms.}\end{aligned}$$

Hence, $R = 10.4$ ohms resistance per mile, and $L = 23.4/(6.28 \times 1000) = 0.00373$ henry per mile. From equation 89, $\gamma/Z_O = y = G + j\omega C = 0.0365 \underline{/77.98°}/702 \underline{/-11.9°} = 0.000052 \underline{/89.88°} = 0.000000104 +$

$j0.000052$. Hence, $G = 0.000000104$ mho or 0.104 micromho per mile, and $C = 0.000052/(6.28 \times 1000) = 0.00000000828$ farad or 0.00828 microfarad per mile.

The calculations just made were for an artificial line designed to have characteristics similar to those of a non-pole pair side circuit of a 104-mil, hard-drawn copper, open-wire line with 12-inch spacing, the constants of which are listed in Table IV. The open-circuited and short-circuited tests used in the calculations were made on this laboratory artificial line.

REFERENCES

1. *American Standard Definitions of Electrical Terms.* A.I.E.E., 1941.
2. WEINBACH, M. P. *Principles of Transmission in Telephony.* The Macmillan Co.
3. KENNELLY, A. E. *The Application of Hyperbolic Functions to Electrical Engineering Problems.* McGraw-Hill Book Co.
4. JOHNSON, K. S. *Transmission Circuits for Telephonic Communication.* D. Van Nostrand Co.
5. FLEMING, J. A. *The Propagation of Electric Currents in Telephone and Telegraph Conductors.* D. Van Nostrand Co.
6. EVERITT, W. L. *Communication Engineering.* McGraw-Hill Book Co.
7. MCILWAIN, KNOX, and J. G. BRAINERD. *High Frequency Alternating Currents.* Second Edition. John Wiley & Sons.
8. WARE, L. A., and H. R. REED. *Communication Circuits.* John Wiley & Sons.
9. GREEN, E. I. Transmission Characteristics of Open-wire Telephone Lines. *Bell System Tech. J.,* October, 1930, Vol. 9, No. 4.
10. TERMAN, F. E. *Electrical Engineers' Handbook.* McGraw-Hill Book Co.
11. Radio Instruments and Measurements. *Bureau of Standards Circular* No. 74.
12. WILSON, L. T. A Study of Telephone Line Insulators. *Trans. A.I.E.E.,* October, 1930, Vol. 49.
13. HOWARD, B. F. Controlling Insulation Difficulties in the Vicinity of Great Salt Lake. *Jour. A.I.E.E.,* December, 1926, Vol. 45, No. 12.
14. GREEN, E. I. Methods of Measuring Insulation of Telephone Lines at High Frequencies. *Trans. A.I.E.E.,* Vol. 46, 1927.
15. STERBA, E. J., and C. B. FELDMAN. Transmission Lines for Short-wave Radio Systems. *Proc. I.R.E.,* July, 1932, Vol. 20, No. 7.
16. LAPORT, E. A. Open-wire Radio-frequency Transmission Lines. *Proc. I.R.E.,* June, 1943, Vol. 31, No. 6.
17. RAGAZZINI, J. R. Ultrashort Electromagnetic Waves—Transmission Lines at Ultrahigh Frequencies. *Elec. Eng.,* April, 1943, Vol. 62, No. 4.
18. BURROWS, C. R. The Exponential Transmission Line. *Bell System Tech. J.,* October, 1938, Vol. 17, No. 4.
19. WHEELER, H. A. Transmission Lines with Exponential Taper. *Proc. I.R.E.,* January, 1939, Vol. 27, No. 1.
20. BURROWS, C. R. The Exponential Transmission Line. *Bell Lab. Record,* February, 1940, Vol. 8, No. 6.

REVIEW QUESTIONS

1. Distinguish between a *uniform line* and a *non-uniform* line.
2. In communication, what frequencies are transmitted by open-wire lines?
3. What types of distortion may occur in open-wire lines and terminal equipment?
4. How do electrically long lines differ from electrically short lines? Give an example of each. How is each considered in design calculations?
5. *Discuss the linear electrical constants* of a transmission line, explaining the fundamental nature and variations of each constant.
6. Of what significance are the phase constant β and the attenuation constant α?
7. Define, and explain, the practical significance of characteristic impedance. What comparable term was considered in the preceding chapter?
8. Why is the input impedance of a line with any termination different from the characteristic impedance of that line?
9. What is the nature of the characteristic impedance of any open-circuited or short-circuited line?
10. Briefly discuss the importance of the reflected wave in determining transmission-line phenomena.
11. What is meant by the term stationary, or standing, wave? Explain why this term is misleading.
12. What are antinodes, and what are nodes? How are they caused? How are they measured?
13. What causes skin effect? Proximity effect?
14. Is skin effect of practical importance? Where? To what extent?
15. Discuss the nature of the losses occurring in telephone insulators, insulator pins, and crossarms.
16. Could a measurement using equation 71 be used to determine the conductance (G) values of Table IV? If not, how would they be determined?
17. Compare the constants of an open-wire line at audio frequencies with those of the same line at radio frequencies.
18. Compare α, β, and Z_0 for a line at audio frequencies, and for the same line at radio frequencies.
19. What is an exponential transmission line, and what interesting property does it have? It is similar to what network of the preceding chapter?
20. What is meant by radiation loss as given by equation 81?
21. What is a phantom circuit? A repeating coil?
22. Explain how, at radio frequencies, a transmission line can be used as a transformer.
23. How would you determine the characteristic impedance of an audio-frequency line? Of a radio-frequency line?
24. How would you determine line constants at both audio and radio frequencies?
25. What is the approximate velocity of propagation of a typical audio-frequency line? What is the attenuation? Repeat for a typical radio-frequency line.

PROBLEMS

1. Calculate the resistance, inductance, and capacitance for a 128-mil hard-drawn copper, open-wire line, with 12-inch spacing, and at 1000 cycles and 20° centigrade. Compare these values with those given in Table IV.

2. Calculate α, β, and Z_0 for the line of Problem 1, and at a frequency of 1000 cycles.
3. If the line of Problem 1 is 100 miles long, is terminated in its characteristic impedance, and the input testing power is one milliwatt at 1000 cycles, calculate the input voltage and current, and the received voltage, current, and power.
4. Repeat the calculations starting on page 227 at a frequency of 15 megacycles.
5. Repeat the calculations starting on page 228 at a frequency of 15 megacycles.

CHAPTER 7

CABLES AND WAVE GUIDES

Introduction. From the electrical standpoint alone, open-wire transmission lines compete favorably with cables, including the coaxial type. From the mechanical standpoint, the open-wire line, consisting as it does of exposed uninsulated wires supported on glass insulators, crossarms, and poles, has serious shortcomings. The open-wire line is subject to mechanical contact with trees, brush, and other wires. The insulators become damaged by natural causes and by man. Furthermore, open-wire lines are subjected to severe mechanical stresses due to sleet, snow, and high winds. The wires must be spaced at least approximately 8 inches apart or they will swing together. The maximum number of *pairs* of wires that can exist on a modern pole line is approximately 40; that is, 8 crossarms of 10 wires each.

For these and other reasons, cables must be used extensively in modern communication systems. Cables used in communication are of two general types: *first,* those in which the transmission circuits are twisted pairs of insulated wires, and *second,* those in which the transmission circuit consists of an outer tubular conductor and a conductor held at the center (called a **coaxial cable**).

Cables are sometimes supported by steel strands attached to poles and are called **aerial cables.**[1] There is a growing tendency, however, to bury cables so that they will be more nearly damage and storm proof. These are called **underground cables.**[1]

Cables of the twisted-pair type are extensively used in telephony for transmission at audio frequencies. They are also used for transmission at telephone carrier frequencies (page 429). Coaxial cables are used in telephony for transmitting many telephone conversations simultaneously (page 432) and for television and radio purposes.

Cables of Paired Twisted Wires. Annealed copper wires are used in cables. Each wire is insulated with a paper tape, or with a layer of "paper" applied to the wires in the form of paper pulp.[2] The wires are twisted together into pairs. In some cables used for toll service, two pairs are twisted together into a four-wire arrangement called a **quad**[1] which is used as a **phantom group** (page 224).

The cables are composed of wires of several different A.W.G. sizes. The smallest wire sizes used are 28 gauge.[3] Usually, however, the

smallest wire size used is 26 gauge. With these it is possible to place 4242 wires (2121 talking pairs) in a cable sheath having an outside diameter of $2\frac{5}{8}$ inches. In the past, the cable sheaths were of lead-antimony alloy. This sheath excludes moisture and protects the wires mechanically. Cables with plastic sheaths* are also used.

Linear Electrical Constants of Cables. The linear electrical constants of cables are the series resistance R, the series self-inductance L, the shunt capacitance C, and the shunt conductance G. In discussing these constants, two types of cables must be considered: those used for local or exchange purposes and those used for toll and long-distance purposes. Also, the cables must be considered at audio frequencies and at the higher carrier frequencies.

These so-called constants are of such a nature that they should be called **parameters.**[4] They are determined from measurements on actual short cable sections.[4] Much of the following information is from reference 4.

Series Resistance R. Both skin effect (page 215) and proximity effect cause the alternating-current or effective resistance to be greater than the direct-current resistance. **Proximity effect** is defined[1] as "the phenomenon of non-uniform current distribution over the cross-section of a conductor caused by variation of the current in a neighboring conductor."

The effective resistance of a cable pair at various frequencies can be computed by means of the curves of Fig. 1. For instance, a 19-gauge conductor at 20° centigrade has a direct-current resistance of about 8 ohms per 1000 feet. At a carrier frequency of 51,200 cycles, the corresponding value of B is $B = \sqrt{51{,}200/8} = 80$. If skin effect only is to be considered, the effective resistance would be about 1.12 times the direct-current resistance, or about 8.96 ohms per 1000 feet. If the proximity effect *of the other wire of the pair* is to be considered and if k of Fig. 1 is 0.4 (which is about the ratio actually measured in a cable[4]), then the effective resistance at this same frequency is about 1.3 times the direct-current resistance, or about 10.4 ohms per 1000 feet. Proximity effects of wires other than the companion wire of the pair have been neglected in Fig. 1.

The variations in the effective resistance of a cable pair with changes in temperature are computed[4] as outlined on page 217. Values of the effective resistance for 19-gauge cable pairs, which are widely used in toll and long-distance circuits, are shown in Fig. 2.

* See "Alpeth Cable Sheath" by R. P. Ashbaugh, *Bell Lab. Record*, November, 1948, Vol. 26, No. 11.

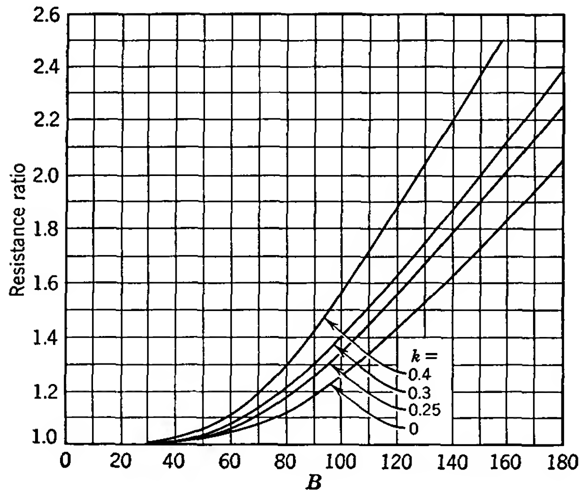


FIG. 1. Curves for determining the series resistance of telephone cable conductors when skin effect only is considered ($k = 0$), and when both skin effect and proximity effect ($k = 0.25, 0.3, 0.4$) are considered. The factor k is the ratio of wire radius to distance between centers, expressed in the same units. For a typical telephone cable, $k = 0.4$. The value of B is the square root of the ratio of the frequency in cycles, divided by the direct-current resistance in ohms, per 1000 feet *per wire*. The direct-current resistance is multiplied by the resistance ratio to obtain the alternating-current resistance (see text). (Reference 4.)

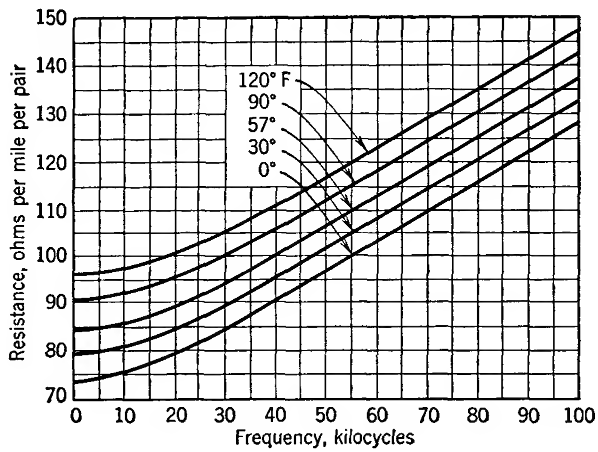


FIG. 2. Series resistance per mile for both wires of a 19-gauge telephone cable pair at various frequencies and temperatures. (Reference 4.)

Series Inductance L. The series self-inductance of two parallel wires is given by equation 69, page 217. This equation contains a term $\mu\delta$ that corrects for the skin effect. Another term should be added to

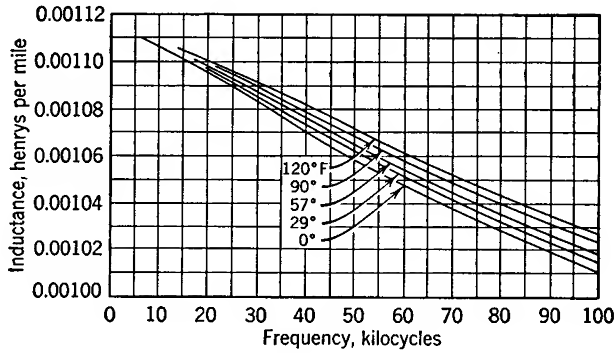


FIG. 3. Series self-inductance per mile for both wires of a 19-gauge telephone cable pair at various frequencies and temperatures. (Reference 4.)

this equation to correct for the proximity effect. For engineering purposes, however, it is advisable to obtain inductance values for cables from sources such as Table I, page 250, Table III, page 253, or from Fig. 3.

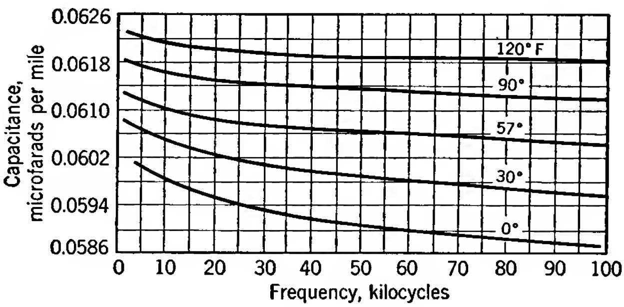


FIG. 4. Shunt capacitance per mile for 19-gauge telephone cable pairs at various frequencies and temperatures. (Reference 4.)

Shunt Capacitance C. The capacitance between two parallel wires is given by equation 70, page 218. This equation applies to the two wires in free space and does not apply with accuracy to twisted pairs surrounded by other wires and in a conducting lead sheath. Also, this formula is for an air dielectric, but the dielectric constant of the paper insulation used in cable pairs is from 1.7 to 1.9, depending on the

amount of air and impurities contained in the paper.[4] Frequency and temperature affect the dielectric constant in a complicated way,[4] and thus again, for engineering purposes, it is advisable to obtain the values of cable capacitance from tables such as Table I, page 250, Table III, page 253, or from Fig. 4.

Shunt Conductance G. The exact nature of the losses that determine the shunt conductance of a cable pair is quite involved, perhaps even more so than for an open-wire line (page 219). Cables offer the advantage, however, that, with the exception of temperature effects, weather conditions have negligible influence on the shunt conductance. The nature and moisture content of the dielectric, the frequency of the

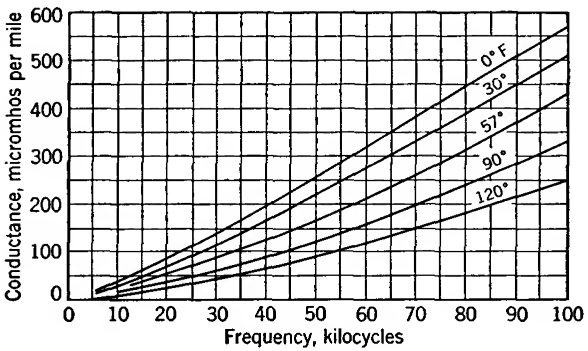


Fig. 5. Shunt conductance per mile for 19-gauge telephone cable pairs at various frequencies and temperatures. (Reference 4.)

test voltage, the spacing of the wires, and the wire sizes are factors affecting the shunt conductance.[4] When manufactured, cables are carefully dried to reduce conductance. The shunt conductance of cables is given in Table I, page 250, Table III, page 253, and in Fig. 5.

The data presented in this section have been for R, L, C, and G of twisted-pair cables to 100,000 cycles. Such cables have been studied[4] at higher frequencies, although at present they are seldom used at these higher frequencies.

Propagation Constant and Characteristic Impedance of a Cable. The transmission of an electromagnetic signal wave over a two-wire cable pair follows the theory of the preceding chapter. For instance, a cable pair should be terminated in its characteristic impedance to avoid wave reflection phenomena. An examination of the constants of cables, given in Tables I and III, will disclose that both the series inductance L and the shunt conductance G are so low, compared to the

series resistance R and shunt capacitance C, that the inductance and conductance may be neglected for approximate calculations.

Approximate Propagation Constant of a Cable. The general expression for the propagation constant is given by equation 36, page 200, as $\gamma = \sqrt{(R + jX)(G + jB)}$. When the substitutions $X = \omega L$ and $B = \omega C$ are made and L and G are assumed to be zero, then

$$\gamma = \sqrt{j\omega RC} = \sqrt{\omega RC}\ \underline{/45^\circ}. \tag{1}$$

From this equation, or from equations 48 and 49 of page 201, when L and G are assumed zero,

$$\beta = \sqrt{\frac{\omega RC}{2}}, \tag{2}$$

and

$$\alpha = \sqrt{\frac{\omega RC}{2}}, \tag{3}$$

where β will be in radians per mile and α will be in nepers per mile when R is in ohms per loop mile, C is in farads per mile of cable, and ω is 2π times the frequency in cycles per second.

Approximate Characteristic Impedance of a Cable. The characteristic impedance of any two-wire transmission circuit is given by equation 50, page 202, as $Z_O = \sqrt{(R + j\omega L)/(G + j\omega C)}$. If L and G are negligible,

$$Z_O = \sqrt{\frac{R}{j\omega C}} = \sqrt{\frac{R}{\omega C}}\ \underline{/-45^\circ}. \tag{4}$$

The magnitude of the characteristic impedance will be in ohms when R is in ohms, C is in farads, and ω is 2π times the frequency in cycles per second. The values of R and C can be for any lengths, provided that they are the same. Since the characteristic impedance of a circuit determines the magnitude and phase angle of the current, it is important to note that from equation 4 the current at each point along a cable will lead the voltage at that point by 45°. This assumes, of course, that the cable is terminated in its characteristic impedance.

Actual values of β, α, and Z_O will be found in Table I, page 250, Table III, page 253, and the high-frequency loss in decibels in Fig. 6. Additional data at high frequencies are given in reference 4.

The Distortionless Line. About 1890, Heaviside showed theoretically that, if the linear line constants (or parameters) were so related that $LG = CR$, transmission conditions would be improved.

If this relation among the *distributed* constants exists, then equation 48, page 201, simplifies to

$$\beta = \omega\sqrt{LC}, \quad \text{and} \quad V = \frac{\omega}{\beta} = \frac{1}{\sqrt{LC}}, \tag{5}$$

and this velocity has been shown (page 226) to be the velocity of light

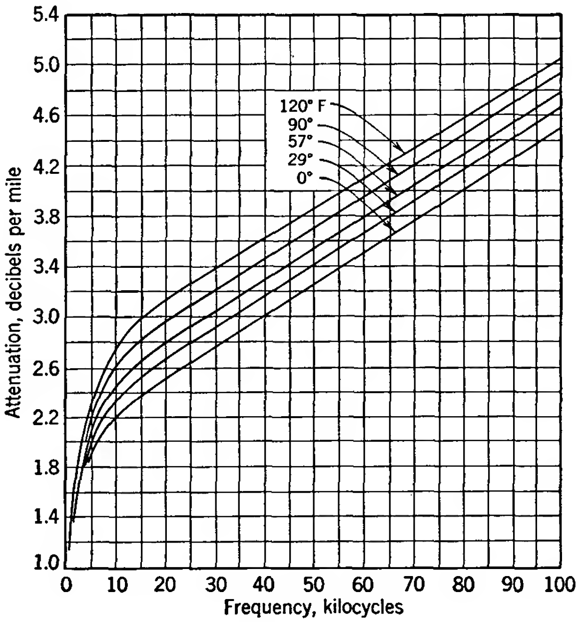


FIG. 6. Attenuation of 19-gauge telephone cable pairs at various frequencies and temperatures. (Reference 4.)

in free space. Also, if $LG = CR$, then equation 49, page 201, becomes

$$\alpha = \sqrt{RG}, \tag{6}$$

and also, equation 50, page 202, becomes

$$Z_O = \sqrt{R/G} = \sqrt{L/C}\,. \tag{7}$$

Derivations of these equations will be found in references 5, 6, 7, and 8.

Thus, if $LG = CR$, the wave velocity, the attenuation, and the characteristic impedance are independent of frequency. If complex voltage waves, such as of speech, are impressed on such a circuit, the current entering the line (determined by Z_O) will be independent of the frequency, the rate at which the various components travel along

the line will be independent of frequency, and the attenuation will be independent of frequency. Thus a complex wave will be transmitted without a change in wave form (without distortion), and a transmission circuit in which $LG = CR$ is known as a **distortionless line.**

Other characteristics of the distortionless line are that the attenuation becomes the least possible value, and the characteristic impedance becomes a value of pure resistance and is increased to a high value.

Development of Loading. In his *Electromagnetic Theory,* published in 1893, Heaviside considers[9] on page 441 "various ways, good and bad, of increasing the inductance of circuits." Attempts to follow his suggestions and those of other early investigators did not prove successful. Vaschy also is credited[10, 11] with early contributions. Thomson proposed[7] placing inductive shunts across cable conductors and dividing the line into sections, each section being connected to the adjacent ones by transformers.

Pupin successfully solved the problem of what is called **inductive loading.** He suggested adding series inductance, in the form of carefully constructed coils, at regular and comparatively short intervals.

At about this same time,[14] Campbell, working independently, also developed a theory of inductive loading. Priority was adjudged Pupin, and his patent rights were acquired by the Bell System,[15] whose engineers are largely responsible for the application of loading to modern communication.

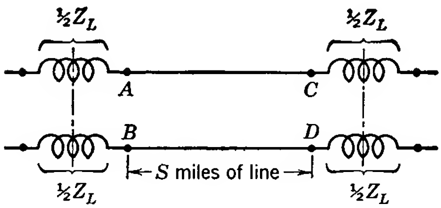


FIG. 7. Method of inserting loading coils in a cable pair. The coils at each loading point are inductively coupled.

Loading was first applied to telephone open-wire lines in about 1900. By the use of loading, talking distances were approximately doubled. By 1925, most loading in the United States had been removed from *open-wire lines.* This was because of the development of the vacuum-tube amplifier. Loading is expensive, and on open-wire lines is particularly susceptible to impairment and damage by heavy transient currents, such as are induced sometimes by lightning. At the present time, *only cables are loaded* in the United States.

Transmission Equations for Coil-Loaded Circuits. In practice, loading coils having an impedance of $\frac{1}{2}Z_L$ are installed in series with *each* cable conductor at spacings S miles apart as indicated in Fig. 7.

A section of this loaded circuit may be represented by the network of Fig. 8, where the part *A–B*, *C–D* is the equivalent T section of the section of cable S (without the loading) having a characteristic impedance Z_O and propagation constant γ per mile.

It is seen from Fig. 8 that the characteristics of the final loaded cable are a combination of the section *A–B*, *C–D* having distributed constants and the lumped loading coils $\frac{1}{2}Z_L$ at each end. The propagation constant γ_L and the characteristic impedance Z_{OL} of the final

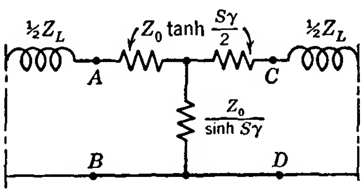


FIG. 8. Network equivalent to Fig. 7.

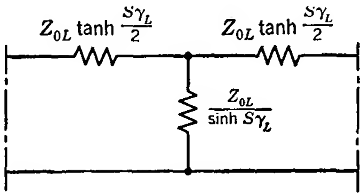


FIG. 9. The T section of a uniform line that is equivalent to the loaded circuit of Figs. 7 and 8.

loaded cable may be considered the propagation constant and the characteristic impedance of an exactly equivalent uniform cable represented diagrammatically by the T section of Fig. 9.

Adding the loading coils to the circuit does not appreciably change the shunt paths of Figs. 8 and 9. It can therefore be written that

$$\frac{Z_O}{\sinh S\gamma} = \frac{Z_{OL}}{\sinh S\gamma_L}, \quad \text{and} \quad Z_{OL} = Z_O \frac{\sinh S\gamma_L}{\sinh S\gamma}. \tag{8}$$

Again referring to these figures, it is seen that

$$\tfrac{1}{2}Z_L + Z_O \tanh \frac{S\gamma}{2} = Z_{OL} \tanh \frac{S\gamma_L}{2}. \tag{9}$$

Then,

$$\tanh \frac{S\gamma_L}{2} = \frac{Z_L}{2Z_{OL}} + \frac{Z_O}{Z_{OL}} \tanh \frac{S\gamma}{2}. \tag{10}$$

When the value of Z_{OL} given by equation 8 is substituted in equation 10,

$$\cosh S\gamma_L = \cosh S\gamma + \frac{Z_L}{2Z_O}, \sinh S\gamma \tag{11}$$

and

$$\gamma_L = \frac{1}{S} \cosh^{-1} \left[\cosh S\gamma + \frac{Z_L}{2Z_O} \sinh S\gamma \right]. \tag{12}$$

This is the Campbell formula[5] for a loaded cable, giving the propagation constant γ_L of the loaded cable in terms of the values of the unloaded cable and the impedance of the loading coils. Considerable mathematical reduction is required between equations 10 and 11; this will be found on page 181 of reference 6.

As shown in reference 6, equation 11 can be written (since $\cosh^2 x = \sinh^2 x + 1$),

$$1 + \sinh^2 S\gamma_L = \cosh^2 S\gamma + \frac{Z_L}{Z_O} \sinh S\gamma \cosh S\gamma + \frac{Z_L^2}{4Z_O^2} \sinh^2 S\gamma, \quad (13)$$

and this can be reduced to the form

$$\sinh S\gamma_L = \sinh S\gamma \sqrt{1 + \frac{Z_L^2}{4Z_O^2} + \frac{Z_L}{Z_O} \coth S\gamma}. \quad (14)$$

When equation 14 is substituted in equation 8, it follows that

$$Z_{OL} = \sqrt{Z_O^2 + \frac{Z_L^2}{4} + Z_O Z_L \coth S\gamma}. \quad (15)$$

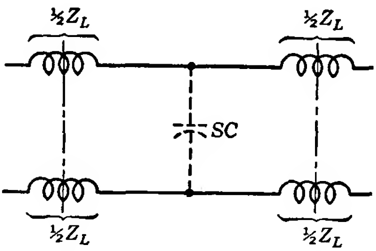


FIG. 10. A coil-loaded cable is equivalent to a low-pass filter.

This formula gives the mid-load or mid-coil characteristic impedance of the lump-loaded cable in terms of the constants of the cable, before loading, and the impedance of the loading coils. A similar equation can be derived for the midsection termination.

Cutoff Frequency of Coil-Loaded Circuits. In cable circuits the inductance is low and the capacitance is high. When loading coils are installed in such circuits at spaced intervals of S miles, the final loaded cable may be represented by Fig. 10, which is essentially that of a low-pass filter.

If the losses are neglected, from equation 70, page 168, the circuit will cut off at a frequency of

$$f_c = \frac{1}{\pi\sqrt{LSC}} \text{ (approximately).} \quad (16)$$

When L is the final self-inductance in henrys in *each loading section* (approximately the sum of the inductance added to *each* wire by the coils), S is the spacing between coils in miles, and C is the capacitance in farads per mile; f_c will be the cutoff frequency in cycles per second.

The various equations derived for low-pass filters also give approximate solutions for loaded cable circuits. Thus, equation 73, page 169, becomes

$$Z_O = \sqrt{L/SC}. \tag{17}$$

In loading a cable of C farads capacitance per mile, there are two variables: these are the *inductance* of the coils and the *spacing*, and these determine the cutoff frequency and final characteristic impedance of the cable.

For the given cutoff frequency and coil inductance, the spacing S in miles, for a cable of C farads per mile, is, from equation 16

$$S = \frac{1}{\pi^2 f_c^2 LC}, \tag{18}$$

and from equation 17 in terms of the characteristic impedance

$$S = \frac{L}{Z_O^2 C}. \tag{19}$$

If equations 18 and 19 are equated,

$$L = \frac{Z_O}{\pi f_c}, \tag{20}$$

and if this is placed in equation 18,

$$S = \frac{1}{\pi f_c C Z_O}. \tag{21}$$

Transmission Characteristics of Coil-Loaded Cables. Cables are loaded to approach the relation $LG = CR$ by inserting inductance, in the form of ferromagnetic-cored coils, in series with the wires at regular intervals. The desired relation could be achieved by altering other constants. However, when all factors are carefully studied, the conclusion is reached that adding inductance is the most satisfactory means. The condition $LG = CR$ is approached but not reached; among the reasons for this are the facts that adding inductance in the form of coils, called **lump loading,** also adds effective resistance, and also that a heavily coil-loaded circuit (one to which much inductance has been added) has a low velocity of propagation (Table III) and tends to accentuate echo effects (page 409).

A comparison of the attenuation losses in a non-loaded and a loaded cable circuit is shown in Fig. 11. In the non-loaded circuit the loss is about *four times* as great at 2000 as at 200 cycles, and considerable frequency distortion would result. The loss is not only much lower

in the loaded circuit but also fairly constant until the vicinity of the cutoff frequency. The "lumpiness effect" is due[11] "to repeated internal reflections at points of electrical discontinuity in the line caused by the insertion of the loading coils."

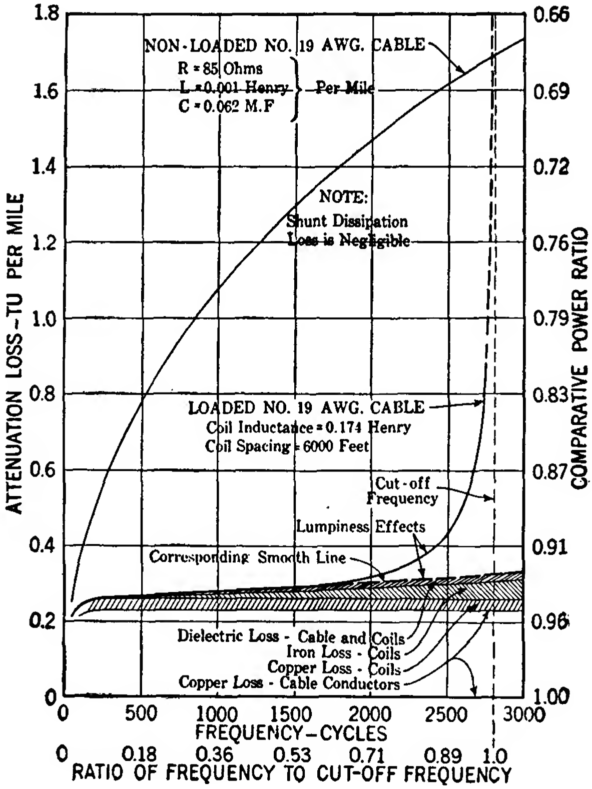


FIG. 11. Attenuation-frequency characteristics of loaded and non-loaded 19-gauge cable pairs. Loss in transmission units (TU), now termed decibels. (Reference 11.)

The velocity of propagation for the non-loaded and loaded cable is shown in Fig. 12. As indicated, for the non-loaded circuit the wave velocity at 2000 cycles is more than three times as great as at 200 cycles, thereby causing delay distortion. It is important to note that, although the wave velocity of the particular loaded circuit shown is fairly constant, it is very low, a condition accentuating echo effects.

Impedance of Coil-Loaded Cables. The variations with frequency of the characteristic impedance of non-loaded and loaded toll cable circuits are shown in Fig. 13. The characteristic impedance

of the non-loaded circuit is low and reactive, being approximately $Z_O = \sqrt{\frac{R}{\omega C}} \underline{/-45^\circ}$ (equation 4). For the loaded circuit, however, $Z_O = \sqrt{L/C}$ (equation 7) and is almost entirely resistive.

As is evident from Fig. 13, the characteristic impedance of a loaded cable circuit depends upon the type of termination used. The **mid-**

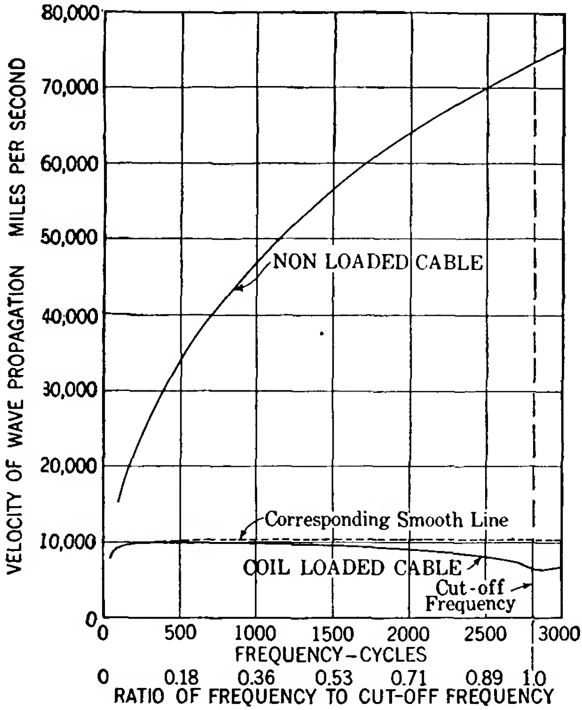


FIG. 12. Velocity-frequency characteristics of loaded and non-loaded 19-gauge cable pairs. (Reference 11.)

section termination is shown in Fig. 14, and, as indicated, the first full loading coil is located at one-half the length of a loading section from the end of the line. This is accordingly equivalent to a π low-pass filter, and the characteristic impedance of a loaded cable with midsection terminations is similar to that of the π filter section (Fig. 28, page 170).

For the **midcoil,** or **midload, termination** of Fig. 15, the two coils placed at the two ends of the circuit have one-half the inductance of the regular coils. This arrangement is similar to the T low-pass filter,

and thus the characteristic impedance is similar to that of such a filter.

Combinations of fractional loading sections and fractional coils are frequently used, especially at the junction of open-wire lines and loaded

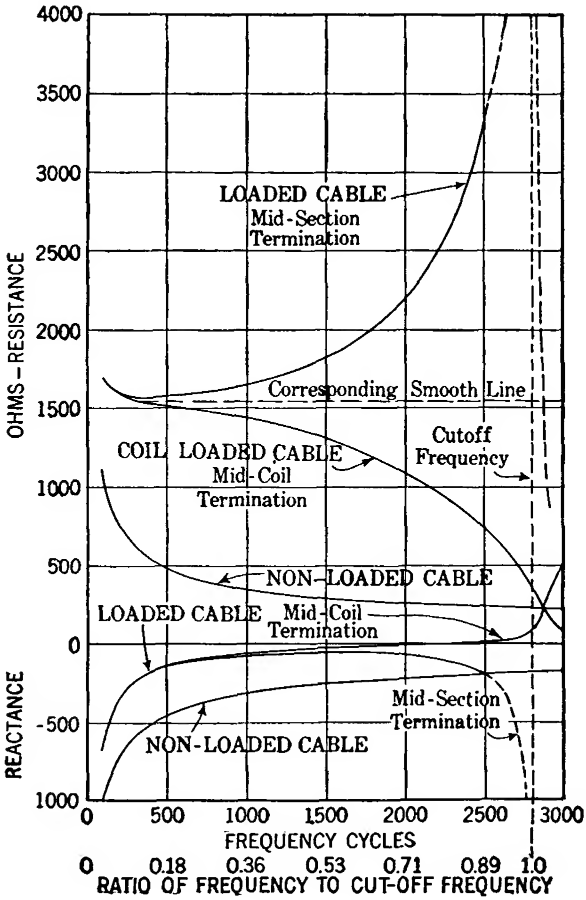


FIG. 13. Impedance-frequency characteristics of loaded and non-loaded 19-gauge cable pairs. (Reference 11.)

cables, to reduce the impedance mismatch. This is similar to the use of m-type filter sections (page 182).

Tables I, II, and III give the characteristics of cables. Of particular interest are the cable circuits for program transmission[16] listed last in Table III. Data for both long toll cables and local or exchange cables are included. These latter are usually unloaded, except in the case of the trunks connecting the various central offices, and particularly long local circuits.

Loading Coils.[17, 18, 19, 20] Circuits not arranged for phantom service (page 225) are loaded with coils such as those shown in Fig. 16. These consist of two windings separated by fiber washers represented

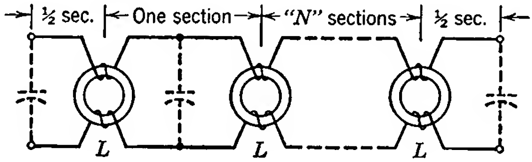


FIG. 14. Midsection termination of a loaded cable.

by the heavy black lines. The magnetic cores are continuous, although cores with air gaps have been used.[11] Since the instantaneous cur-

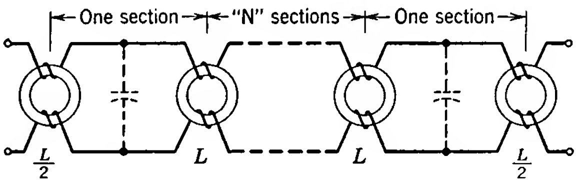


FIG. 15. Midcoil or midload termination of a loaded cable.

rents in the two line wires are in opposite directions, the flux produced by each winding adds to that of the other, producing the maximum inductive effect.

The first loading coils had continuous ring-shaped cores consisting of a coil of many turns of fine iron wire. The wires were insulated to reduce eddy-current losses. In 1916 an important improvement was made when a core of powdered and compressed iron was perfected. A core of powdered and compressed Permalloy was developed[17] about 1926. For about the same efficiency, the Permalloy core gives a coil very much smaller than the iron-core type, and hence it permits reductions in size. A further reduction in the size of loading coils is made practicable by the use of the molybdenum Permalloy dust core shown in Fig. 17. A detailed summary of recent improvements in cable loading coils and a description of the method of winding the coils is given in reference 20.

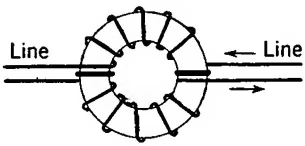


FIG. 16. Loading coil used on non-phantomed circuits.

Loading Phantom Group Circuits. Although loading was developed about 1900 and soon was applied to the side circuits, it was

FIG. 17. Equivalent cores for loading coils. Iron dust core (left); Permalloy dust core (center); molybdenum Permalloy dust core (right). (Courtesy Bell Telephone System.)

not until about 1910 that phantom circuits were successfully loaded.[10, 11] One reason for this delay was the difficulty of providing the close balance required in the coil windings to prevent excessive crosstalk. A complete loading system for a phantom group is shown

TABLE I

CONSTANTS OF EXCHANGE PAPER-INSULATED CABLES, COPPER CONDUCTORS

Gauge	Resistance R, Ohms per Loop mile 68° F	Capacitance C, Microfarads per Mile	Leakance G, Micromhos per Mile 1000 Cycles	Characteristic Impedance Magnitude, Ohms 1000 Cycles
26 (BST)	440	0.079	2.1	$942\angle{-44.5°}$
24 (DSM)	274	0.084	2.2	$721\angle{-44.2°}$
22	171	0.082	2.1	$576\angle{-44°}$
19	85	0.084	2.2	$402\angle{-43°}$
19 (Special)	85	0.066	1.7	$453\angle{-43°}$

NOTE: Inductance is 0.001 henry per loop mile for all gauges.

in Fig. 18. The theory is illustrated by the simplified circuit of Fig. 19, where the broken arrows represent instantaneous currents and corresponding fluxes produced by the phantom circuit, and the continuous arrows the currents and resulting fluxes produced by the currents in the side circuits.

The currents in the side circuits produce adding fluxes, and the maximum inductance is therefore obtained for loading the side circuits. In the phantom loading coil, however, these currents are in opposite directions through identical windings, and thus *neither side circuit will produce magnetic flux in the core of the phantom loading coil.* Similarly, the phantom-circuit currents pass through the separate side-circuit

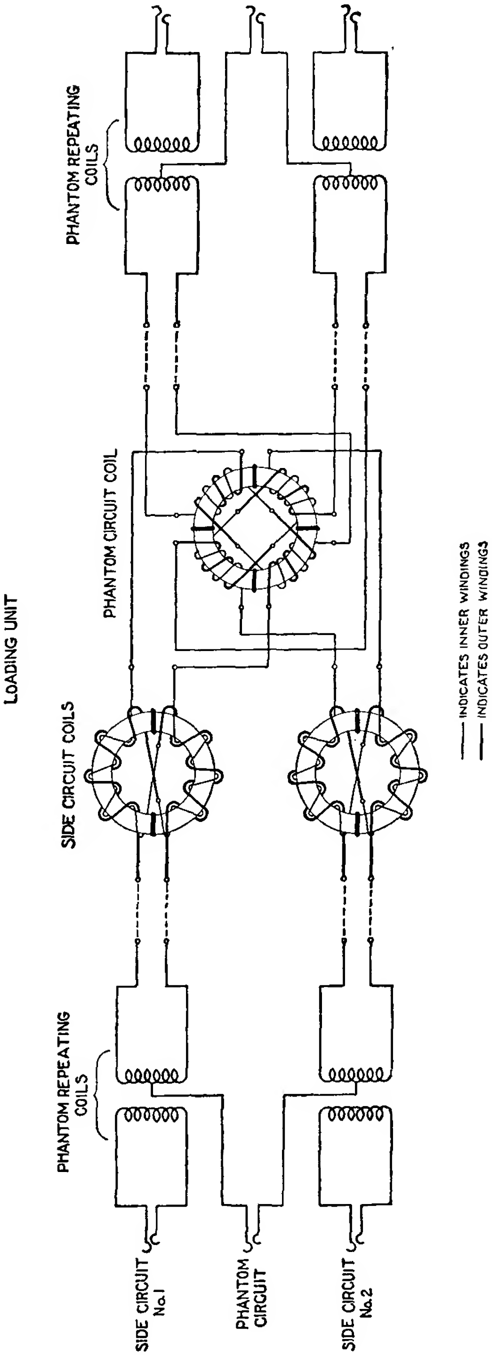


FIG. 18. Standard method of loading phantom and side cable circuits in the Bell System.

TABLE II

CHARACTERISTICS OF EXCHANGE CABLE CIRCUITS

Circuit Constants at 1000 Cycles—68° F					Nominal Cutoff Cycles per Second
Cable*	Loading†	Characteristic Impedance‡	Propagation Constant per Mile Nepers Radians	Attenuation, Decibels per Mile	
26-BST	NL	$942\underline{/-44.5°}$	0.3278 + j0.3322	2.86	....
	H-88	$1192\underline{/-20.8°}$	.2076 + j .5403	1.80	
24-DSM	NL	$721\underline{/-44.2°}$	.2664 + j .2715	2.31	....
	H-88	$1074\underline{/-14.6°}$	.1414 + j .5341	1.23	....
	B-88	$1416\underline{/-\ 8.1°}$	.1078 + j .7298	0.94	....
22-CSA	NL	$576\underline{/-43.8°}$	.2065 + j .2134	1.79	....
	M-88	$905\underline{/-13.7°}$	.1060 + j .4341	0.92	2860
	H-88	$1051\underline{/-\ 9.7°}$	.0907 + j .5185	0.79	3510
	H-135	$1306\underline{/-\ 6.3°}$	.0729 + j .6402	0.63	2820
	B-88	$1420\underline{/-\ 5.3°}$	.0689 + j .718	0.60	4960
	B-135	$1765\underline{/-\ 3.3°}$	.0549 + j .890	0.48	3990
19-CNB	NL	$402\underline{/-42.8°}$	.1446 + j .1551	1.26	....
	M-88	$861\underline{/-\ 9.4°}$	.0568 + j .4302	0.49	2830
	H-88	$1017\underline{/-\ 5.2°}$	.0487 + j .5194	0.42	3460
	H-135	$1283\underline{/-\ 3.3°}$	.0388 + j .6455	0.34	2790
	B-88	$1395\underline{/-\ 2.8°}$	.0386 + j .725	0.34	4900
	B-135	$1742\underline{/-\ 1.7°}$	.0304 + j .900	0.26	3950
19-DNB	NL	$453\underline{/-42.8°}$	.1282 + j .1375	1.11	....
	M-88	$956\underline{/-\ 7.4°}$	.0505 + j .3796	0.44	3190
	H-88	$1137\underline{/-\ 5.2°}$	.0432 + j .4590	0.38	3910
	H-135	$1425\underline{/-\ 3.3°}$	.0345 + j .5694	0.30	3150
	B-88	$1565\underline{/-\ 2.8°}$	.0342 + j .641	0.30	5520
	B-135	$1952\underline{/-\ 1.8°}$	.0270 + j .795	0.24	4450

* Paper-insulated cables. See Table I for primary constants.

† Numerals indicate total inductance of loading coils in millihenrys, letters indicate load coil spacing as follows: NL, non-loaded; M, 9000 feet; H, 6000 feet; B, 3000 feet.

‡ Midsection iterative impedance in cases of loaded circuits. Magnitude in ohms. Although M-88 and H-135 loading is in service, it is no longer used in new installations.

coils in the same direction, and since these windings are identical, *no magnetic flux is produced in the side-circuit coils by the phantom-circuit current.* Under these conditions of balance there is no cross-talk due to magnetic coupling.

TABLE III

Characteristics of Typical Paper-insulated Toll Cable Circuits

Copper Conductors—1000 cycles per second

Type of Circuit	Wire Size AWG	Type of Loading	Spacing of Load Coils, Miles	Load Coil Constants per Load Section: R Ohms D-C	Load Coil Constants per Load Section: L Henrys	Constants Assumed to be Distributed per Loop mile: R Ohms A-C	Constants Assumed to be Distributed per Loop mile: L Henrys	Constants Assumed to be Distributed per Loop mile: C Microfarads	Constants Assumed to be Distributed per Loop mile: G Micromhos	Propagation Constant per Mile, Polar: Magnitude Ohms	Propagation Constant per Mile, Polar: Angle (Positive)	Propagation Constant per Mile, Rectangular: α Nepers	Propagation Constant per Mile, Rectangular: β Radians	Midsection Characteristic Impedance, Polar: Magnitude Ohms	Midsection Characteristic Impedance, Polar: Angle (Negative)	Midsection Characteristic Impedance, Rectangular: R Ohms	Midsection Characteristic Impedance, Rectangular: X Ohms (Negative)	Velocity, Miles per Second	Nominal Cut-off Frequency Cycles	Attenuation Decibels per Mile 55° F
Side.....	19	NL	.....	.....	....	83.5	0.001	0.062	1.4	0.181	47° 28′	0.122	0.133	469	42° 36′	345	317	47,240	....	1.06
Phantom	19	NL	.....	.....	....	41.7	.0007	.100	2.2	.162	49° 14′	.106	.123	259	40° 36′	197	169	52,400	....	0.92
Side.....	19	H-172-S	1.136	13.8	0.172	95.7	.152	.062	1.4	.642	87° 12′	.0313	.640	1660	2° 48′	1658	82	9,800	2,800	.27
Phantom	19	H-63-P	1.136	6.9	.064	47.8	.056	.100	2.2	.481	86° 10′	.0321	.478	782	3° 42′	789	50	13,100	3,700	.28
Side.....	19	H-88-S	1.136	7.9	.089	90.5	.079	.062	1.4	.447	84° 49′	.0403	.445	1160	5° 12′	1155	105	14,100	4,000	.35
Phantom	19	H-50-P	1.136	4.0	.050	45.3	.044	.100	2.2	.428	85° 28′	.0338	.427	687	4° 36′	685	55	14,700	4,200	.30
Side.....	19	H-44-S	1.136	4.5	.044	87.5	.039	.062	1.4	.330	80° 35′	.0541	.326	821	9° 24′	810	135	19,300	5,600	.47
Phantom	19	H-25-P	1.136	2.2	.025	43.7	.022	.100	2.2	.310	81° 40′	.0449	.306	494	8° 12′	489	70	20,600	5,900	.39
Side.....	19	B-88-S	0.568	7.9	.089	97.4	.157	.062	1.4	.623	87° 6′	.0316	.622	*1543	*3° 18′	*1540	*90	10,100	5,600	.28
Phantom	19	B-50-P	0.568	4.0	.050	48.8	.088	.100	2.2	.597	87° 27′	.0265	.596	*921	*2° 29′	*920	*40	10,600	5,900	.23
Side.....	16	NL	.....	.....	....	41.1	.001	.062	1.4	.122	49° 7′	.0796	.0920	332	40° 6′	254	214	66,840	....	.69
Phantom	16	NL	.....	.....	....	20.5	.0007	.100	2.2	.115	50° 59′	.0721	.0890	186	36° 30′	150	111	70,600	....	.61
Side.....	16	H-172-S	1.136	13.8	.172	53.3	.152	.062	1.4	.641	88° 20′	.0186	.640	1658	1° 36′	1658	45	9,800	2,800	.16
Phantom	16	H-63-P	1.136	6.9	.064	26.6	.056	.100	2.2	.479	87° 47′	.0185	.478	781	2° 36′	780	35	13,100	3,700	.16
Side.....	16	H-88-S	1.136	7.9	.089	48.1	.079	.062	1.4	.438	87° 3′	.0224	.437	1155	2° 29′	1155	50	14,100	4,000	.19
Phantom	16	H-50-P	1.136	4.0	.050	24.0	.044	.100	2.2	.423	87° 26′	.0189	.422	686	2° 31′	685	30	14,700	4,200	.16
Side.....	16	H-44-S	1.136	4.5	.044	45.1	.039	.062	1.4	.322	84° 54′	.0286	.320	813	5° 00′	810	71	19,600	5,600	.25
Phantom	16	H-25-P	1.136	2.2	.025	22.4	.022	.100	2.2	.305	85° 31′	.0238	.304	483	4° 23′	482	37	20,700	5,900	.21
Side.....	16	B-88-S	0.568	7.9	.089	55.0	.157	.062	1.4	.619	88° 17′	.0185	.618	*1541	*1° 29′	*1540	*40	10,100	5,600	.16
Phantom	16	B-50-P	0.568	4.0	.050	27.6	.088	.100	2.2	.594	88° 29′	.0157	.593	*920	*1° 15′	*920	*20	10,600	5,900	.14
Physical .	16	B-22-N	0.568	0.9	.022	42.6	.039	.062	1.4	.317	85° 3′	.0273	.315	*803	*4° 39′	*800	*65	19,600	11,000	.24

* Note: Midcoil characteristic impedance.

What is called **flutter** may occur in loaded circuits simultaneously providing telephone and telegraph facilities.[21] The relatively low-frequency telegraph currents passing through the windings cause corresponding changes in the effective resistance and the inductance of the coils. These produce variations in the transmitting efficiency of the coils, and thus the talking currents vary and produce a "fluttering" effect.

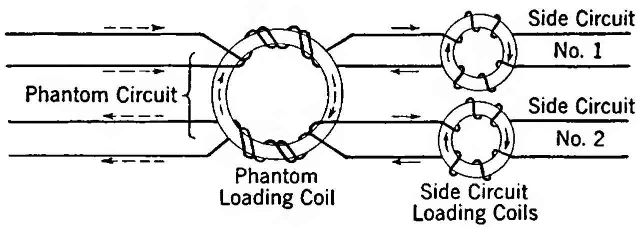


FIG. 19. Broken arrows represent current and flux due to phantom. Full arrows represent currents and flux due to side circuits.

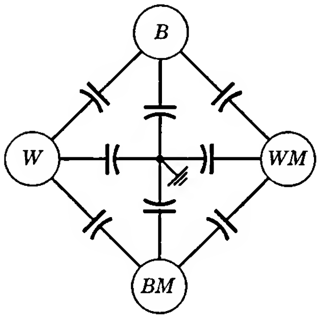


FIG. 20. End view of a cable quad showing distributed capacitances.

The Installation of Cable Loading. When a cable is loaded, extensive "capacity unbalance" tests are made and the unbalances minimized.[22] Figure 20 represents the end view of a cable "quad" or the four wires of a phantom group. Each wire has capacitance to each other wire and to ground (cable sheath), and these can be represented as shown. If each of the corresponding capacitances is the same, then no capacitive unbalance exists, and no cross-talk will be produced from this source. That is, a conversation on the "white" pair *W* and *WM* (white and white mate) will not be heard on the "black" pair; neither will a conversation on the phantom *B–BM* and *W–WM* be transferred to either side circuit *B* and *BM* or *W* and *WM*.

Cables are designed and manufactured to keep these unbalances small. When a cable is installed in service, adjacent lengths are spliced together so that the unbalances in one length tend to neutralize those in the next.[22, 23]

The Location of Impedance Irregularities. In loading cable circuits it is possible to reverse the loading coil connections and install the coil so that the magnetic effects oppose, and thus the coil will have an

incorrect inductance. Such a discontinuity and other **impedance irregularities** in a line or cable (such as those caused by high-resistance joints, defective loading coils, or transposition errors in open wire and splicing errors in cables) cause wave reflections, and thus the measured impedance characteristics of a circuit may vary as in Fig. 21. Such

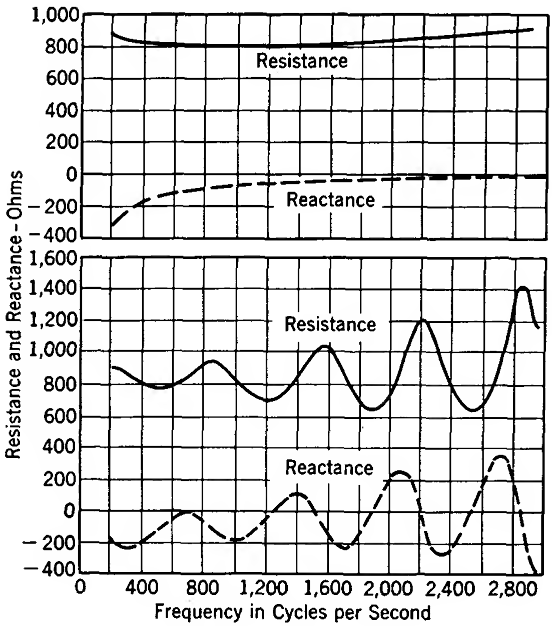


FIG. 21. Characteristics of a 16-gauge paper-insulated cable side circuit, with H-44-S loading. The line without irregularities is shown above. The impedance characteristics of the line with an irregularity is shown below. (Data from supplement to *Bell System Tech. J.*, July, 1936, Vol. 15, No. 3.)

irregularities interfere with telephone operation if they do not altogether prevent it.

It will be observed that cable irregularity causes periodic variations in both the resistance and the reactance components. It is possible from such curves to locate line impedance irregularities.

When an electromagnetic wave strikes an impedance irregularity, part of the wave is reflected. The magnitude of the reflected wave depends on the nature of the unbalance existing. When this reflected wave reaches the sending end, it may either *aid* or *oppose* the current entering the line (depending on the frequency and distance to the unbalance) and thus makes the measured sending-end impedance

($Z = E/I$) either *higher* or *lower* as Fig. 22 indicates. If the distance to the irregularity is great and the velocity of propagation is low, many reflected wavelengths will be included, and hence only a small frequency change is required to add another wavelength between the source and the irregularity.

The equation for determining the distance to the irregularity can be derived by referring to Fig. 22. In this figure the reflected current subtracts from the current which would exist if the line were smooth (that is, had no irregularity), and thus the sending current is smaller

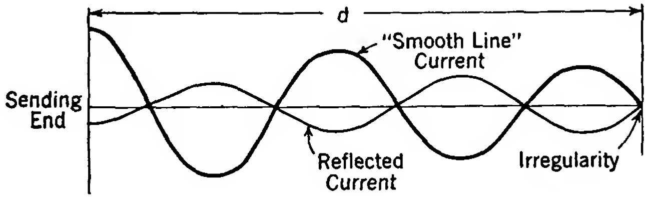


FIG. 22.

and the impedance $Z = E/I$ is maximum. If now the frequency is increased so that, instead of two and one-quarter waves, two and three-quarters waves are included in the distance d, the same reflected current phase relations and a second maximum impedance value will be obtained. Where f_1 and λ_1 are the frequency and wavelength for the first condition, and n is the number of wavelengths included, it can be written that the distance d to the irregularity is

$$d = n\lambda_1. \tag{22}$$

Similarly, for the second condition it can be written that

$$d = (n + \tfrac{1}{2})\lambda_2. \tag{23}$$

The value $n = d/\lambda_1$ obtained from equation 22 can be substituted in equation 23, which then becomes

$$d = \left(\frac{d}{\lambda_1} + \frac{1}{2}\right)\lambda_2 \quad \text{or} \quad d = \left(\frac{d\lambda_2}{\lambda_1} + \frac{\lambda_2}{2}\right). \tag{24}$$

If V is the velocity of propagation in the circuit, then $\lambda = V/f$, and equation 24 becomes

$$d = \frac{df_1}{f_2} + \frac{V}{2f_2}. \tag{25}$$

This may be written

$$d = \frac{V}{2(f_2 - f_1)}, \tag{26}$$

where d is the distance to the irregularity in miles, when V is the velocity of propagation in the circuit in miles per second, and f_1 and f_2 are the frequencies between which impedance peaks occur as in Fig. 21.

The velocity of propagation of the cable of Fig. 21 is about 19,570 miles per second at 1000 cycles. The length between impedance peaks at about this frequency is approximately 650 cycles. Thus from equation 26, $d = 19{,}570/(2 \times 650) = 15.05$ miles (approximately). That is, the irregularity causing the reflections is located about 15 miles from the testing end. During such tests it is, of course, necessary to have the distant end of the line terminated in the line characteristic impedance so that reflections will not be caused at this point.

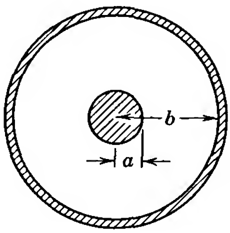


FIG. 23. A coaxial cable consists of a conductor (often solid as shown) at the center of a conducting sheath. The dimensions a and b are usually as indicated.

Coaxial Cables. Coaxial cables differ radically from the conventional telephone cable composed of twisted pairs of insulated wires. Coaxial cables consist of a wire conductor held at the center of a conducting tube or sheath (Fig. 23).

Coaxial cables are used for the simultaneous transmission of many telephone messages, for transmitting television programs, for connecting radio transmitters to the sending antennas, and for connecting radio-receiving antennas to the receiving sets. Since the beginning of communication, coaxial cables have been used for transoceanic telegraph circuits.

The construction details of a coaxial cable depend on the purpose to which it will be put. The common features are the copper conductor held at the center of the outer copper sheath by insulators of hard rubber, ceramic, plastic, or other suitable material. The sheath acts as one side of the transmitting circuit and also as an effective shield. Because of skin effect, the high-frequency signal current being transmitted travels along the inside of the sheath and the unwanted currents induced by extraneous sources travel along the outside of the sheath, which is usually at ground potential.

Coaxial Cable Transmission Equations. In deriving the fundamental transmission equations for the coaxial structure the general method followed in the preceding pages for two parallel wires may be

used. From Fleming's work,[7] at the high frequencies of present interest the propagation constant for the coaxial cable becomes

$$\gamma = \alpha + j\beta = \frac{R}{2Z_O} + \frac{GZ_O}{2} + j\frac{2\pi}{\lambda}. \tag{27}$$

The characteristic impedance at high frequencies is almost pure resistance and equals[24]

$$Z_O = \sqrt{L/C}, \tag{28}$$

where R, G, L, and C are the resistance, shunt conductance, inductance, and shunt capacitance in ohms, mhos, henrys, and farads, respectively, for any convenient length, and at the frequencies under consideration. The factor λ is the corresponding wavelength.

The difficulty in the use of these equations is that the constants vary, at least to some extent, with frequency because of skin effect. This is particularly true for the losses represented by R and G. If certain reasonable assumptions are made, however, simple relations for the coaxial cable at high frequencies can be written.

If the insulation losses given by the central term of equation 27 are neglected, the attenuation constant, giving the loss in an air-dielectric coaxial cable in nepers, is

$$\alpha = R/2Z_O, \tag{29}$$

where Z_O is as given by equation 28. The length unit of measure of α will be the same as for R. The value of the phase constant β in radians per meter becomes

$$\beta = 2\pi f/c, \tag{30}$$

where c equals the velocity of light in free space in meters per second (page 195).

The value of the resistance R in ohms per *centimeter* is[24]

$$R = 41.6\sqrt{f}\left(\frac{1}{a} + \frac{1}{b}\right) \times 10^{-9}, \tag{31}$$

where b is the inner radius of the outer conductor in centimeters, a is the outer radius of the inner conductor in centimeters, and f is the frequency in cycles per second. The value of the inductance L is

$$L = 0.46 \log_{10} b/a, \tag{32}$$

where L is in microhenrys per meter and b and a are as shown in Fig. 23 and are measured in the same units. The value of the capacitance C is

$$C = \frac{0.0000241}{\log_{10} b/a}, \tag{33}$$

where C is in microfarads per meter and b and a are as previously considered. When the last two equations are substituted in equation 28, the characteristic impedance in ohms becomes

$$Z_O = 138 \log_{10} b/a \tag{34}$$

and substantially equals pure resistance.

Coaxial Cables Used in Telephony. A photograph of the original New York–Philadelphia coaxial cable, installed about 1936, is shown in Fig. 24. With this cable and a special carrier system, it is theoretically possible to have 240 simultaneous telephone conversations. With

FIG. 24. View of the original coaxial cable installed between New York and Philadelphia. The ordinary wires are for voice-frequency circuits for talking purposes during tests. (Courtesy Bell Telephone System.)

later coaxial cables (page 434) and associated carrier equipment (page 432) it is possible to have *600 simultaneous telephone conversations.* As many as eight coaxial conductors are used in coaxial telephone **composite cables,** which also contain ordinary cable pairs in addition to the coaxial conductors. The various geographical regions of the United States are now interconnected with a coaxial cable system.

Coaxial Cables Used in Radio. Coaxial cables used in radio are of two general types: those with air dielectrics (and ceramic or other insulators) and those with continuous, or solid, dielectrics, such as polyethylene. The attenuation characteristics of the air-dielectric cables can be found from the equations given. The characteristics of certain coaxial cables with solid (yet flexible) dielectrics are given in Fig. 25.

It was early shown[24] that if the ratio $b/a = 3.6$, then the attenuation of a coaxial cable with air dielectric is, theoretically, a minimum value. From equation 34 this ratio gives a characteristic impedance of about 77 ohms (resistance). Until about 1940, most coaxial cables were constructed with this ratio, but since that time there has been an increasing tendency to design cables having a characteristic impedance of about 52 ohms.

Coaxial cables used for feeding a broadcast antenna must sometimes transmit many kilowatts.

The following calculations serve to illustrate the use of the preceding equations. A copper coaxial cable having the dimensions $a = 1.59$ centimeters and $b = 3.81$ centimeters is to deliver 20 kilowatts at 1.19

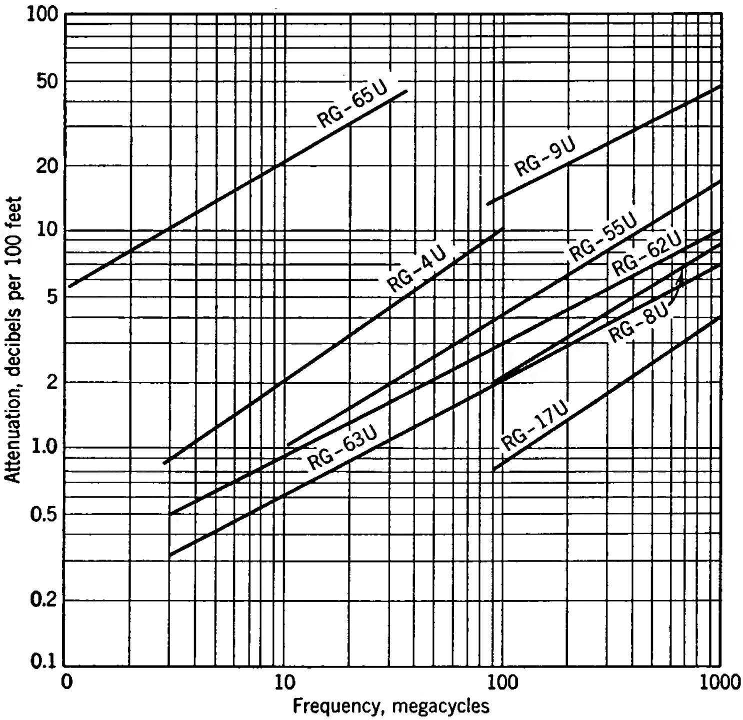


FIG. 25. Attenuation at various frequencies of coaxial cables used in radio. Such cables are in general small, flexible, and insulated with polyethylene. (For further information consult *Reference Data for Radio Engineers*, Federal Telephone and Radio Corp., from which these data were obtained.)

megacycles to an antenna that is matched to the cable so that no standing waves exist. The cable length is 500 feet, or 152.3 meters.

The inductance is found by equation 32 to be $L = 0.46 \log_{10} (3.81/1.59) = 0.175$ microhenry per meter.

The capacitance is found by equation 33 to be $C = 0.0000241/\log_{10} (3.81/1.59) = 0.0000634$ microfarad per meter.

The characteristic impedance is found from equation 28 to be $Z_O = \sqrt{0.175 \times 10^{-6}/0.0000634 \times 10^{-6}} = 52.5$ ohms. Or, from equation 29, $Z_O = 138 \log_{10}(3.81/1.59) = 52.5$ ohms.

The resistance is found from equation 31 to be $R = 41.6\sqrt{1.19 \times 10^6} \times \left(\frac{1}{1.59} + \frac{1}{3.81}\right) \times 10^{-9} = 40.4 \times 10^{-6}$ ohm per centimeter.

The attenuation is found from equation 29 to be $\alpha = 40.4 \times 10^{-6}/2 \times 52.5 = 0.385 \times 10^{-6}$ neper per centimeter $= 0.385 \times 10^{-6} \times 8.686 = 3.34 \times 10^{-6}$ decibel per centimeter, or 0.0508 decibel for the 500-foot length of coaxial cable.

The voltage at the distant end will be $E = \sqrt{PZ_O} = \sqrt{20 \times 10^3 \times 52.5} = 1025$ volts.

The current at the distant end will be $I = \sqrt{P/Z_O} = \sqrt{20 \times 10^3/52.5} = 19.5$ amperes.

The power input at the sending end will be $P = 20 \times 10^{0.1 \times 0.0508} \doteq 20{,}200$ watts.

The voltage at the sending end will be $E = \sqrt{PZ_O} = \sqrt{20.2 \times 10^3 \times 5.25} = 1030$ volts.

The current at the sending end will be $I = \sqrt{P/Z_O} = \sqrt{20.2 \times 10^3/52.5} = 19.6$ amperes.

The efficiency of transmission will be $20/20.2 = 0.99$ or 99 per cent.

Coaxial Cables as Circuit Elements. Wave-reflection phenomena were considered in the preceding chapter; corresponding phenomena may exist on coaxial cables. An open-wire line is a *balanced* unshielded transmission circuit, and a coaxial cable is an *unbalanced* shielded circuit; in other respects they are fundamentally the same.

From this it follows that the explanations, diagrams and equations of the preceding chapter apply in general to coaxial cables. Sections of coaxial cables, therefore, can be used as inductors, capacitors, transformers, etc.

Submarine Telegraph Cables. The first successful telegraph cable was laid across the Atlantic in about 1865. In general telegraph cables are coaxial, and the equations and discussions previously given for these cables apply. Actually, however, such cables were operated at such low telegraphic speeds that they were usually designed on a direct-current basis.

Transoceanic submarine telegraph cables were first inductively loaded in 1924. Although coil or lumped loading is used for short submarine cables in relatively shallow water, for mechanical reasons it is not well adapted for long transoceanic submarine cables. Krarup or **continuous loading** is accordingly used for such cables.[25]

Cables are continuously loaded by wrapping the conductors with either a magnetic tape or wire. This increases the number of magnetic lines of force encircling the wire for a given current value and thus

uniformly increases the inductance. The fundamental transmission equations of Chapter 6 for circuits with distributed constants apply. A continuously loaded cable does not exhibit the filter action present in coil-loaded circuits. The first loaded submarine telegraph cables used Permalloy tape, closely wound on the central copper conductor (Fig. 26).

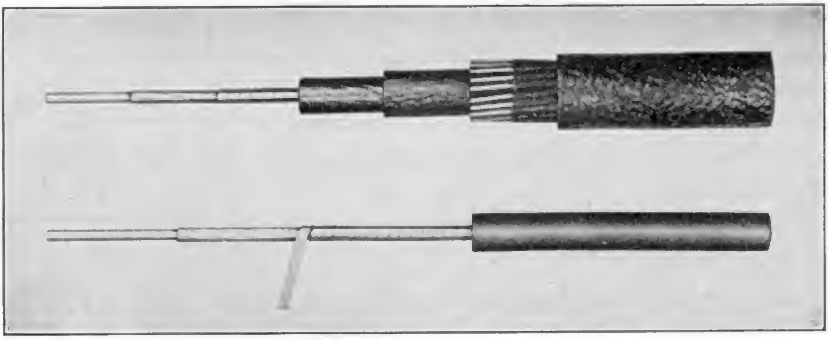


FIG. 26. Permalloy-loaded cable. Above, section of deep-sea type from New York-Horta cable. Below, section of core showing Permalloy tape partly unwound. (Reference 25.)

Submarine Telephone Cables. Such cables are used across rivers, bays, etc., for *telephone* purposes. If the distance is but a few miles, the cables are armored cables of the conventional twisted-pair type. If the cables must be longer and if the water is deep, special coaxial cables are used.

Telephone service was established initially[26] between the United States and Cuba over three continuously loaded cables laid between Key West and Havana and opened for service in 1921. They are slightly over 100 miles long and are laid in water more than a mile deep.

The three original Key West–Havana telephone cables were continuously loaded with a layer of iron wire. In later cables Permalloy was used. Not all submarine telephone cables, however, are loaded. Carrier systems are sometimes used on submarine cables, and in such instances[27] the loading may be omitted because of excessive carrier-frequency losses.

Important long submarine telephone cables have been installed from Italy to Sardinia,[28] from Australia to Tasmania,[29] and elsewhere. No transoceanic *telephone* cables have been laid, however, the word transoceanic implying a major ocean such as the Atlantic.

A submarine transatlantic telephone cable was designed about 1930 and *was* to be laid between Newfoundland and Ireland, about 2100 miles, but was not installed. No intermediate amplifiers were to be used. By the use of continuous Perminvar (page 62) loading and Paragutta insulation the overall loss and distortion were to be maintained within practical limits.[30]

Studies have been made[31] of the possibilities of a transatlantic submarine cable on which perhaps 12 telephone channels could operate. It has been proposed that the cable system be provided with underwater vacuum-tube amplifiers (or repeaters) at numerous points and that these be *built into the cable structure.*[32]

Wave Guides. The transmission lines and cables considered in the preceding pages are, in a sense, wave guides. They *guide* the electromagnetic waves that contain the electric energy transmitted. Lines and cables are seldom treated as guides, however.

A solid dielectric material such as wood, glass, or polystyrene will guide electromagnetic waves. These materials have dielectric constants greater than air, and electromagnetic waves will follow a stick or rod of such material in preference to air, but the attenuation is high.

A *hollow* metal tube of circular, rectangular, or other cross-sectional shape, also will guide energy, in the form of electromagnetic waves, from one point to another. Such hollow metal tubes are what is usually meant by the term **wave guide.** Wave guides are used at ultrahigh and superhigh frequencies (page 442) largely because the radiation loss and attenuation loss in open-wire lines and coaxial cables become excessive at such frequencies.

The electromagnetic wave that transmits the electric energy is established at the sending end of the wave guide by an arrangement of one or more conductors that is similar to a radio antenna. Once the energy is transferred to the hollow wave guide, it will be transmitted along the guide until the energy is dissipated in the walls of the guide, or absorbed by other means; the high-frequency energy cannot escape through the conducting walls. In fact, the currents that are established in the walls of the wave guide by the passing electromagnetic waves flow so completely in the surface layers that silver plating the inner surfaces materially reduces the transmission losses. At the distant end of the wave guide the received electromagnetic energy is extracted by an arrangement of conductors constituting a receiving antenna, or the energy is radiated by an **electromagnetic horn.**

From the engineering viewpoint, it is possible to study wave guides by analogy with transmission-line theory. Wave guides are at present (1949) of limited practical importance, relatively speaking, in com-

mercial communication systems. The use of wave guides in commercial communication systems undoubtedly will grow. Perhaps it is not absurd to envision a transcontinental wave-guide system. References 8 and 33 to 36 are recommended for information on wave guides.

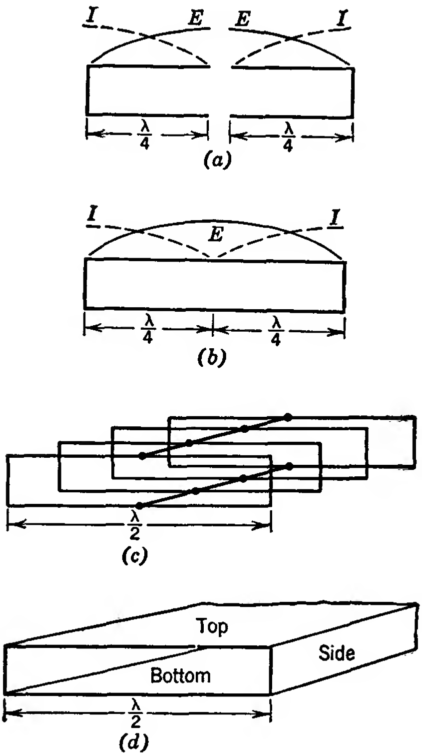


FIG. 27. Diagrams for illustrating certain elementary relations between short-circuited transmission lines and wave guides.

Rectangular Wave Guides. In Fig. 27(*a*) are shown two shorted quarter-wave sections. If these are assumed to be essentially lossless, the input impedance of each section will be *very* high. If these two sections are electrically excited by connecting a source of high-frequency voltage across the open end, by inducing high-frequency currents in the shorted end, or by other convenient means, the current and voltage distribution along the quarter-wave lines will be as shown.

If the two wires are joined as in Fig. 27(*b*) and excited as previously explained, the voltage and current distribution remain essentially the same. If several such conductors are connected as in Fig. 27(*c*), the voltage and current will again be distributed as in (*b*). If now, many shorted quarter-wave conductors are placed side by side in perfect contact, the rectangular wave guide of Fig. 27(*d*) results.

If the rectangular wave guide of Fig. 28 is excited (by a vertical "antenna" wire passing through the center of the guide) at a frequency such that the greatest cross-section dimension is $\lambda/2$ (as in Fig. 27), it follows that the voltage and current distribution is as indicated. If this is true, within the wave guide the electric-field component and the magnetic-field component of the electromagnetic wave must be as shown in Fig. 28(*a*). At the center of the guide the voltage is large as indicated by the height of curve E; hence, the electric field will be

intense as shown by the concentration of vertical solid arrows. The voltage across the "short circuit" caused by the sidewalls of the tube is

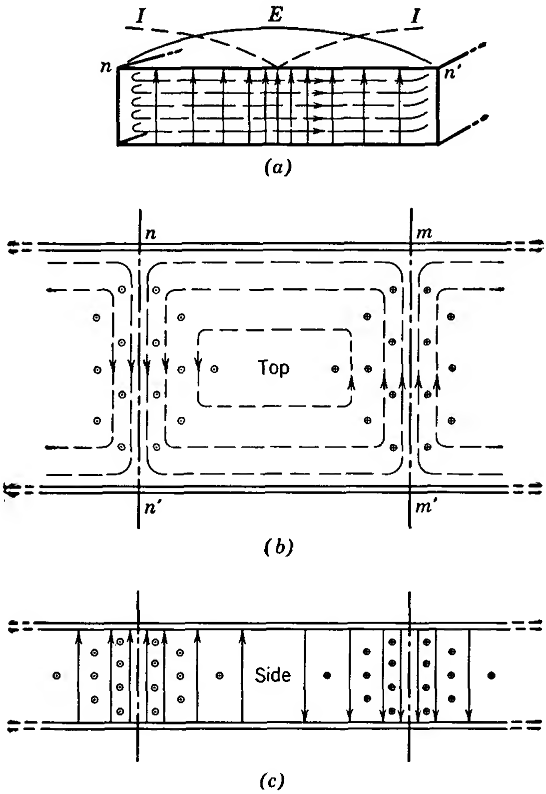


FIG. 28. Approximate field distributions in rectangular wave guides for the $TE_{1,0}$ mode. In (*a*), the solid arrows are the electric field and the broken arrows are the magnetic field, looking into the wave guide at transverse section n–n'. Energy flow is *in*. The magnetic lines are, of course, continuous. In (*b*) is shown the magnetic field distribution, looking into the top of the guide. The electric field is shown, as dots at section n–n' and crosses at m–m'. Energy flow is to the right. In (*c*) is shown the field distribution, looking into the side of the guide. The arrows represent the electric field, and the dots and crosses represent the magnetic field. Energy flow is to the right.

zero (by analogy with a transmission line), and hence no electric field exists at these walls.

The magnitude of the current flowing in the *inner surfaces* of the wave guide is shown by curve I of Fig. 28(a). Large currents are flowing in the sidewalls, and in the top and bottom of the guide near

the sides; hence, the magnetic field will be as shown by the broken horizontal lines. Because magnetic lines of force are continuous, the magnetic lines of Fig. 28(*a*) must be arranged within the guide as shown by the top view of Fig. 28(*b*). A side view of the fields is shown in Fig. 28(*c*).

Cutoff Frequency. A wave guide is similar to a high-pass filter in that there is a low-frequency limit beyond which it will not conduct. If the cross-sectional dimensions of a rectangular wave guide are as shown in Fig. 29, the lowest frequency f_o (and the longest *free-space* wavelength λ_o) that will be transmitted down the guide is given by the equation

$$f_o = \frac{c}{2a} \quad \text{or} \quad \lambda_o = \frac{c}{f_o} = 2a, \tag{35}$$

where f_o is the frequency in cycles per second, λ_o is in centimeters, c is the velocity in centimeters per second of a wave *in free space,* and a is the dimension in centimeters indicated in Fig. 29. From this it is noted that the lowest frequency that will be transmitted is that of a wave that will just fit into a space twice the width of the guide. For instance, if a rectangular wave guide has an a dimension of 7.5 centimeters, it will pass a wave of minimum frequency $f_o = 3 \times 10^{10}/15 = 2 \times 10^9$ cycles or 2000 megacycles, and a wavelength of $\lambda_o = 2 \times 7.5 = 15$ centimeters. Attention is called to the fact that the dimensions a and b are often interchanged. The usage in this book is in accordance with the standards.[37] As is evident the short dimension b has no effect in determining the cutoff frequency. However, a very high voltage (Fig. 28) may exist between the top and the bottom of a guide, and, if dimension b is small, the guide may flash across.

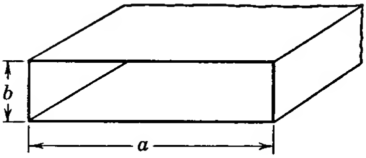


FIG. 29. Dimensions of a rectangular wave guide are specified in the standards as shown.

Mode of Operation. The distribution of electric and magnetic lines of force of Fig. 28 is one of many possible arrangements. This specific arrangement and the values given by equation 35 are for the **dominant mode,** which, for a given wave guide, is the mode of operation that will give the lowest cutoff frequency and the longest wavelength that the guide will transmit. In many applications a wave guide is operated in the vicinity of its dominant mode, one reason being that the attenuation is low for this condition.

The mode of operation of Fig. 28 is known as the $\mathbf{TE}_{1,0}$ mode, for

the following reasons. In free space (page 443) electric energy travels away from the source (antenna) as a **transverse electromagnetic** or **TEM wave.** That is, *both* the electric lines of force and the magnetic lines of force composing the wave are at right angles to the direction of wave propagation. In a wave guide, in which energy is confined to the space within the metal guide, the resultant wave traveling down the guide is composed of two or more waves that are reflected back and forth within the guide. For this reason the resultant wave is not a wave composed of an electric and a magnetic field at right angles to the direction of propagation and is not a TEM wave. In wave guides, there are components of either the electric field or the magnetic field in the direction of propagation. If a wave guide is assumed to be lossless, in the **transverse electric (TE) mode,** there is no electric-field component in the direction of propagation of the resultant wave. Similarly, in the **transverse magnetic (TM) mode,** there is no magnetic-field component in the direction of propagation of the resultant wave.

As mentioned previously the mode of operation in Fig. 28 is known as the $TE_{1,0}$ mode. The meaning of the two capital letters has been explained. For *rectangular* wave guides, the first subscript denotes the number of half waves of electric field intensity (voltage) along the large dimension a, and the second subscript denotes the number of half waves of electric-field intensity (voltage) along the small dimension b. These designations are different from the designations often used but are in accordance with the standards.[37]

Wave Velocities. In considering transmission over open-wire lines and cables, the phase constant, or phase shift per unit distance, was calculated from the line constants, and the wave velocity and wavelength were computed from the phase constant. This is entirely satisfactory for transmission over such circuits, and, because of the way in which the values are found, the term **phase velocity** is often applied. It will also be recalled that wavelength and phase velocity were determined by measuring standing wave peaks on open-wire lines.

But in a wave guide, in a sense, the energy being propagated down the guide is being reflected back and forth within the tube along a zigzag path from wall to wall.[34] If a probe is inserted into a wave guide[33] and if the distribution of the resultant electric field is studied, it will be found that the distance between maximum values or "peaks" of electric field is such that *apparently* the wavelength is longer than would be expected. Since in general it is considered that $\lambda = v/f$ and $v = \lambda f$, it follows that, if the *apparent* wavelength in a guide (as measured with a probe) is greater than it would be in free space, then the **apparent wave velocity** or phase velocity in a guide is greater

than the velocity of an electromagnetic wave in free space. This apparent phase constant β is given by the equation

$$\beta = \sqrt{\left(\frac{\omega}{c}\right)^2 - \left(\frac{\pi}{a}\right)^2}, \tag{36}$$

where β will be the apparent phase shift in radians per centimeter, $\omega = 2\pi f$, c is the velocity of light in centimeters per second, and a is the long dimension of the guide in centimeters (Fig. 29). If β is known, then the apparent or phase velocity in centimeters is

$$V = \frac{\omega}{\beta}. \tag{37}$$

For the wave guide previously considered, and at a frequency of 3×10^9 cycles,

$$\beta = \sqrt{\left(\frac{6.28 \times 3 \times 10^9}{3 \times 10^{10}}\right)^2 - \left(\frac{3.14}{7.5}\right)^2} = 0.47 \text{ radian per centimeter.}$$

Using this value the apparent or phase velocity is

$$V = \omega/\beta = 6.28 \times 3 \times 10^9/0.47 = 4.02 \times 10^{10} \text{ centimeters per second,}$$

which is considerably above the actual velocity that an electromagnetic wave in free space may have, that is, 3×10^{10} centimeters per second. The corresponding apparent wavelength in the guide is $\lambda_g = V/f = 4.02 \times 10^{10}/3 \times 10^9 = 13.88$ centimeters; yet the *actual wavelength in free space* would be $V_s = 3 \times 10^{10}/3 \times 10^9 = 10$ centimeters. Thus, the apparent, or phase, velocity is greater than the velocity in free space and also is greater than the **group velocity,** which is the velocity with which a signal, such as a speech-modulated wave, would be propagated down the guide. These phenomena are caused by the zigzag path previously mentioned.[34]

Attenuation. Attenuation can be computed from equations or can be obtained from curves.[39] For copper walls and an air dielectric and at the dominant or $TE_{1,0}$ mode, the attenuation of a rectangular wave guide is given by the relation[39]

$$\alpha = \frac{0.01107}{a^{1.5}}\left[\frac{\frac{0.5a}{b}\left(\frac{\lambda_c}{\lambda}\right)^{1.5} + \frac{1}{\sqrt{\lambda_c/\lambda}}}{\sqrt{\left(\frac{\lambda_c}{\lambda}\right)^2 - 1}}\right], \tag{38}$$

where α will be in decibels *per foot* when a and b are the dimensions *in*

inches shown in Fig. 29, λ_c is the cutoff wavelength, and λ is the free-space wavelength.

Other Modes of Operation. Other methods of excitation and other modes of operation are used. Furthermore, wave guides other than the rectangular type are employed (references 33 to 36).

REFERENCES

1. *American Standard Definitions of Electrical Terms.* A.I.E.E. 1941.
2. WALKER, H. G., and L. S. FORD. Pulp Insulation for Telephone Cables. *Bell System Tech. J.*, January, 1933, Vol. 12, No. 1.
3. *Bell Lab. Record,* January, 1943, Vol. 21, No. 5.
4. HEBBERT, C. M. The Transmission Characteristics of Toll Telephone Cables at Carrier Frequencies. *Bell System Tech. J.*, July, 1941, Vol. 20, No. 3.
5. JOHNSON, K. S. *Transmission Circuits for Telephonic Communication.* D. Van Nostrand Co.
6. WEINBACH, M. P. *Principles of Transmission in Telephony.* The Macmillan Co.
7. FLEMING, J. A. *The Propagation of Electric Currents in Telephone and Telegraph Conductors.* D. Van Nostrand Co.
8. WARE, L. A., and H. R. REED. *Communication Circuits.* John Wiley & Sons.
9. GILL, F. Oliver Heaviside. *Bell System Tech. J.*, July, 1925, Vol. 4, No. 3.
10. GHERARDI, B. Commercial Loading of Telephone Circuits in the Bell System. *Trans. A.I.E.E.*, Vol. 30, 1911.
11. SHAW, T., and W. FONDILLER. Development and Application of Loading to Telephone Circuits. *Trans. A.I.E.E.*, Vol. 45, 1926.
12. PUPIN, M. I. Propagation of Line Electric Waves. *Trans. A.I.E.E.*, Vol. 16, 1899.
13. PUPIN, M. I. Wave Transmission over Non-Uniform Cables and Long Distance Air Lines. *Trans. A.I.E.E.*, Vol. 17, 1900.
14. *The Collected Papers of George Ashley Campbell.* American Telephone and Telegraph Co. See also *Bell Telephone Quarterly,* January, 1938, Vol. 17, No. 1.
15. RHODES, F. L. Permalloy: The Latest Step in the Evolution of the Loading Coil. *Bell Telephone Quarterly,* October, 1927, Vol. 6, No. 4.
16. CLARK, A. B., and C. W. GREEN. Long Distance Cable Circuits for Program Transmission. *Trans. A.I.E.E.*, October, 1930, Vol. 49. Also *Bell System Tech. J.*, July, 1930, Vol. 9.
17. SHACKELTON, W. J., and I. G. BARBER. Compressed Powdered Permalloy. *Trans. A.I.E.E.*, April, 1928, Vol. 47. Also *Bell System Reprint* B-328, July, 1928.
18. ELMEN, G. W. Magnetic Alloys of Iron, Nickel, and Cobalt. *Elec. Eng.*, December, 1935, Vol. 54, No. 12.
19. LEGG, V. E. Survey of Magnetic Materials and Applications in the Telephone System. *Bell System Tech. J.*, July, 1939, Vol. 18, No. 3.
20. HALE, S. G., A. L. QUINLAN, and J. E. RANGES. Recent Improvements in Loading Apparatus for Telephone Cables. *A.I.E.E. Technical Paper* No. 48–59, December, 1947.
21. GIVEN, F. J. Flutter Effect in Loading Coils. *Bell Lab. Record,* July, 1930, Vol. 8, No. 11.

22. WEAVER, M. A. The Long Struggle against Cable Crosstalk. *Bell Telephone Quarterly,* January 1, 1935, Vol. 14, No. 1.
23. WEAVER, M. A. Balancing Crosstalk in Toll Cables. *Bell Lab. Record,* August, 1935, Vol. 13, No. 12.
24. STERBA, E. J., and C. B. FELDMAN. Transmission Lines for Short-Wave Radio Systems. *Proc. I.R.E.,* July, 1932, Vol. 20, No. 7.
25. BUCKLEY, O. E. High-Speed Ocean Cable Telegraphy. *Bell System Tech. J.,* April, 1928, Vol. 7, No. 2.
26. MARTIN, W. H., G. A. ANDEREGG, and B. W. KENDELL. Key West–Havana Submarine Telephone Cable System. *Trans. A.I.E.E.,* Vol. 41, 1922.
27. GILBERT, J. J. Key West–Havana Cable No. 4. *Bell Lab. Record,* May, 1931, Vol. 9, No. 9.
28. Electrical Communication in 1932. *Elec. Communic.,* January, 1933, Vol. 11, No. 3, p. 117.
29. Australia-Tasmania Cable. *Elec. Communic.,* July, 1936, Vol. 15, No. 1.
30. GHERARDI, B., and F. B. JEWETT. World-Wide Telephony—Its Problems and Future. *Bell System Tech. J.,* October, 1932, Vol. 11, No. 4.
31. BUCKLEY, O. E. Transoceanic Telephone Cables. *Bell Lab. Record,* June, 1942, Vol. 20, No. 10.
32. Repeater for Submarine Telephone Cable. *Bell Lab. Record,* July, 1942, Vol. 20, No. 10.
33. BRAINERD, J. G., G. ROEHLER, H. J. REICH, and L. F. WOODRUFF. *Ultra-High-Frequency Techniques.* D. Van Nostrand Co.
34. SKILLING, H. H. *Fundamentals of Electric Waves.* John Wiley & Sons.
35. SARBACHER, R. I., and W. A. EDSON. *Hyper and Ultra-high Frequency Engineering.* John Wiley & Sons.
36. BRONWELL, A. B., and R. E. BEAM. *Theory and Application of Microwaves.* McGraw-Hill Book Co.
37. *Standards on Radio Wave Propagation—Definitions of Terms Relating to Guided Waves.* I.R.E. 1945.
38. TERMAN, F. E. *Radio Engineers' Handbook.* McGraw-Hill Book Co.
39. TERMAN, F. E. *Radio Engineering.* McGraw-Hill Book Co.

REVIEW QUESTIONS

1. Why is there an increasing tendency to use cables for communication?
2. How are telephone cables constructed? What is a quad?
3. Discuss the variations of R, L, C, and G of a cable with temperature and frequency.
4. Discuss the variations of cable attenuation with temperature and frequency.
5. What are the characteristics of the classical "distortionless line"?
6. Briefly discuss the history of loading.
7. In what way is a loaded cable similar to a filter?
8. How are echo effects and loading related?
9. Compare the characteristic impedances and propagation constants of loaded and non-loaded cables.
10. In general, to what types of circuits is loading applied?
11. Discuss the construction of loading coils.
12. Discuss the installation of loading coils.
13. What is flutter, and what is its cause?

14. The location of impedance irregularities, such as a reversed loading coil, is discussed on page 255. Why not use the methods of page 75?
15. What assumption regarding velocity is made in equation 26?
16. Are coaxial cables balanced or unbalanced? From what important circuits do they differ in this respect?
17. What types of dielectrics are used in coaxial cables?
18. What are the characteristic impedances of typical coaxials? Compare with lines and cables.
19. Describe a typical submarine cable.
20. Describe the type of loading used on transoceanic cables.
21. Are transoceanic cables used for telephone purposes? If so, name one. If not, why are they not used?
22. What is meant by the dominant mode of a wave guide?
23. Distinguish between the TE mode and the TM mode.
24. On page 266 it is stated that for certain conditions a wave guide may flash across. How is it possible for this to occur?
25. How is it possible for the apparent, or phase, velocity in a wave guide to be greater than the velocity of an electromagnetic wave in free space?

PROBLEMS

1. Calculate α, β, and Z_0 at 1000 cycles for a non-loaded 19-gauge telephone cable side circuit and compare the results with the values given in Table III.
2. Calculate α, β, and Z_0 at 1000 cycles for a 19-gauge telephone cable side circuit with H-172-S loading and compare the results with the values given in Table III.
3. A section of the cable of Problem 1 is 100 miles long and is terminated in its characteristic impedance. Standard testing power of 1.0 milliwatt at 1000 cycles is put into the cable at the sending end. Calculate the sending-end current and voltage, and the current, voltage, and power at the receiving end.
4. Make the same calculations as in Problem 3, but for a 100-mile section of the loaded cable of Problem 2.
5. Calculate the cutoff frequency for the circuit of Problem 2.
6. Calculate the cutoff frequency for the last circuit of Table III.
7. As stated on page 259, a ratio of $b/a = 3.6$ gives least attenuation. What would have been the attenuation of the coaxial cable of page 260 if the outer conductor had been such as to give this ratio?
8. What would have been the inductance, capacitance, characteristic impedance, and resistance of the coaxial cable of page 260 if the conditions of Problem 7 applied?
9. Repeat the calculations starting on page 260, but at a frequency of 550 kilocycles.
10. A rectangular wave guide has an a dimension of 5 centimeters and a b dimension of 2 centimeters. What will be the cutoff frequency and the longest free-space wavelength?

CHAPTER 8

ELECTRONIC APPLICATIONS IN COMMUNICATION

Introduction. Modern communication circuits utilize numerous devices such as vacuum tubes, metallic rectifiers, and thermistors. For a complete understanding of the chapters that follow, a review of such devices will be given.

Rectifiers. A **rectifier** is defined[1] as "a device which converts alternating current into unidirectional current by virtue of a characteristic permitting appreciable flow of current in only one direction." A **rectifier unit** is defined[1] as "the rectifier with its essential auxiliaries and the rectifier transformer equipment." In practice, this unit is usually referred to as a rectifier.

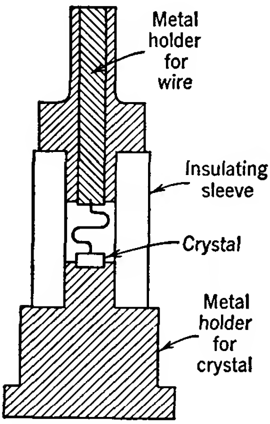


FIG. 1. Simplified cross section of a typical crystal rectifier. Sometimes the interior of a crystal rectifier is filled with wax to stabilize the contact.

Crystal Rectifiers. The cross section of a typical crystal rectifier is shown in Fig. 1. Silicon and germanium crystals are widely used. A fine wire, often of tungsten, is pointed and pressed firmly against the crystal surfaces; sometimes the wire contact is welded to the crystal. Because the capacitance is very low, crystals can be used at ultrahigh and superhigh frequencies (page 442).

The characteristics of a typical crystal are shown in Fig. 2. For many purposes the ideal characteristics and ideal circuit equivalent for a crystal are those shown in Fig. 3. A typical crystal passes a current of about 400 microamperes.

Metallic Rectifiers. Metallic rectifiers consist essentially of a disk or plate of metal on which is formed, or placed, a layer of some suitable semiconducting material. These disks are mounted in series or in parallel, or in series-parallel combinations, until a unit of the desired current and voltage ratings is obtained. They are sometimes called **dry-disk rectifiers** and **barrier-layer rectifiers.** They are of several types, as follows:

Copper—Copper Oxide Rectifiers.[2] Copper oxide rectifiers consist of copper disks or plates on which a layer of cuprous oxide is formed

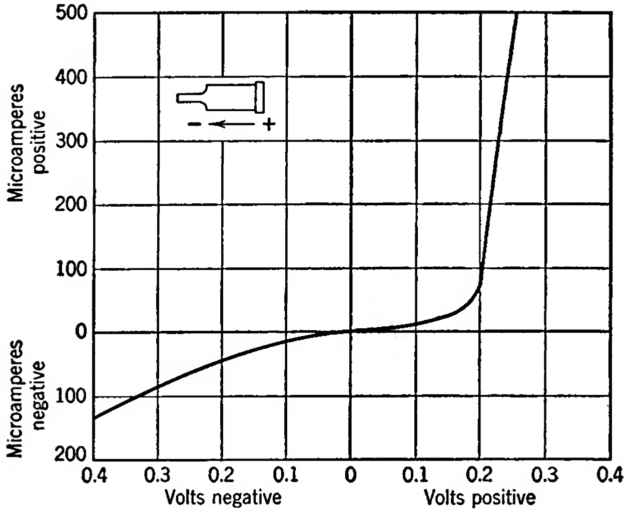


FIG. 2. Characteristics of a crystal rectifier of the type shown in Fig. 1. For "volts positive" the polarities are as shown in the upper left; for "volts negative" they are reversed.

by a controlled heating process.[2, 3] Current (conventional) flows readily from the oxide to the copper, but the resistance is very high in the reverse direction. Considerable capacitance exists at the thin rectifying interface between the copper and the oxide.

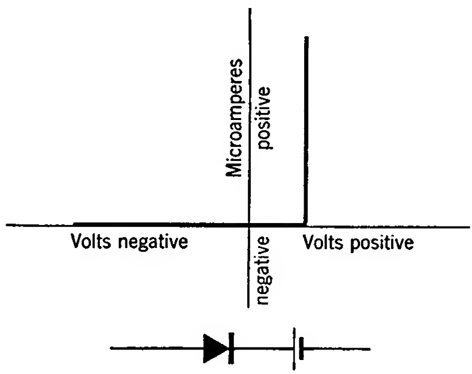


FIG. 3. Idealized characteristic curve and equivalent circuit of the crystal rectifier of Figs. 1 and 2. All such rectifiers do not have exactly the same characteristics.

Various advanced theories have been presented[2] to explain the operation of the **copper oxide rectifier,** as it is commonly called. From a *practical viewpoint* it may be considered that the metallic copper contains many free electrons, and hence, when the oxide is made positive, *electrons* flow readily from the copper to the oxide.

When the polarity is reversed, negligible current flows because the oxide is a poor conductor and contains few free electrons.

Because the cuprous oxide is a poor conductor, a current-distributing electrode must be placed on it. In some rectifiers nickel is electroplated on the surface of the cuprous oxide. Copper oxide rectifiers are often called **varistors,** particularly in communication.

Selenium Rectifiers.[4, 5] Selenium rectifiers consist of a processed metal disk, often of steel, which is called the **back electrode.** On one side of this electrode a very thin layer of selenium is placed, and this layer is covered with a suitable conducting alloy, forming the **front electrode.** A rectifying layer is formed between the front electrode and the selenium. Current readily flows from the back electrode to the front electrode, but little flows in the reverse direction. The characteristics of this device are shown in Fig. 4.

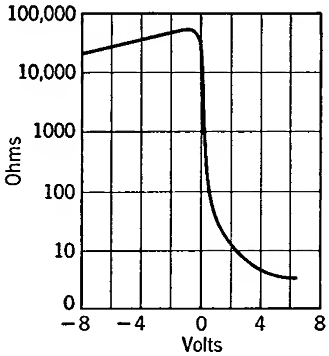


FIG. 4. Approximate characteristics of a selenium rectifier plate one square centimeter in area. Signs refer to voltage of the metal back electrode. (Courtesy *Electrical Communication.*)

High-Vacuum Thermionic Diodes. This high-vacuum diode consists essentially of an electron-emitting hot **cathode** and an electron-collecting **anode** or **plate.** For *small tubes* the cathodes are of two types: first, the **filament type** consisting of nickel wire or some alloy that is coated with oxides of barium and strontium and suitably treated to make an excellent emitter; and second, the **indirectly heated** or **separate-heater type.** This type consists of a small cylinder of nickel or other metal on which the oxide coating is placed and which is heated by an insulated tungsten filament placed at the center. For *large* high-vacuum tubes the filaments are often tungsten.

The emitted **electrons** accumulate as a **negative space charge** about the cathode. If the plate is positive, it pulls negative electrons from the space-charge region. Because the negative space charge repels electrons, it is an important factor in determining the magnitude of the voltage drop across the tube.

The plate is heated by the electrons that strike it, and the **average anode current,** which is the direct current flowing, must not exceed the rated value. The **peak (or crest) inverse anode voltage,** defined[1] as "the maximum instantaneous anode voltage in the direction opposite to that in which the tube is designed to pass current," also must be

kept within rated values. Otherwise, the tube may **arc back** and may be damaged.

Low-Pressure Gas Thermionic Diodes. Gas diodes are used as rectifiers where large currents are to be passed and where the rectified voltages are to be less than about 10,000 volts. These tubes are first evacuated, and then the desired amount of gas, such as argon, is admitted. Or after evacuation a small amount of mercury is placed in the tube. At the low pressure existing, some of the mercury vaporizes, giving the widely used **mercury-vapor tube.**

In the gas or mercury-vapor diode, the space charge around the cathode is neutralized by **positive ions** produced by **ionization by collision** in the following way. If the voltage between the cathode and the anode approximately equals the **ionizing potential** of the gas (about 10 to 12 volts for mercury vapor), then some of the negative electrons being drawn over to the positive anode attain velocities sufficient to knock negative electrons out of the otherwise neutral gas atoms if a collision occurs.

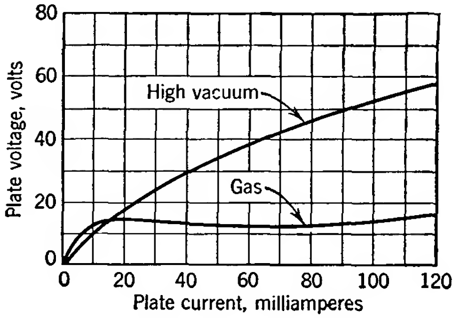


FIG. 5. Characteristics of comparable high-vacuum and gas diodes.

As a result of such collisions, positive ions are formed. They consist of the relatively massive atoms that have lost a negative electron. Because of their large mass, the positive ions drift slowly toward the negative cathode and hence remain in the space between the electrodes for a relatively long time. This accumulation of positive ions neutralizes the negative space charge. For this reason, the voltage drop across a gas or mercury-vapor tube equals approximately the ionizing potential of the gas or vapor. Accordingly, the voltage drop is much lower than that across a high-vacuum tube. Furthermore, over the operating range, the voltage drop is essentially independent of the magnitude of the current being rectified (Fig. 5).

High-Pressure Gas Thermionic Diodes. The trade names **Tungar** and **Rectigon** are applied to high-pressure gas diodes. They are used extensively for charging small storage-battery installations in telephone central offices. The cathodes are massive tungsten or thoriated-tungsten filaments, and they are close to a heavy graphite anode. Argon gas at a pressure of about 5 centimeters of mercury fills the

tube. The gas prevents rapid evaporation of the hot filament and also reduces the space charge.

Mercury-Arc Rectifiers. In a sense, mercury-arc rectifiers are diodes. The cathode is a mercury pool, and it is not externally heated. The mercury-arc rectifier is used for charging telephone central-office batteries.

When the bulb of Fig. 6 is tipped for starting, contact is made between the starting electrode and the mercury-pool cathode. When the bulb is restored to vertical, the contact is broken and the small spark that results fills the tube with positive and negative mercury-vapor ions. Electrons (the negative ions) flow to the anode that is positive at that instant, ionization by collision results, and additional ions are formed. For conduction to continue, electrons must leave the mercury-pool cathode. It is possible that the massive positive ions, slowly moving to the cathode, form a dense "cloud" of positive electricity *very* close to the pool and that this greatly assists in the liberation of electrons.

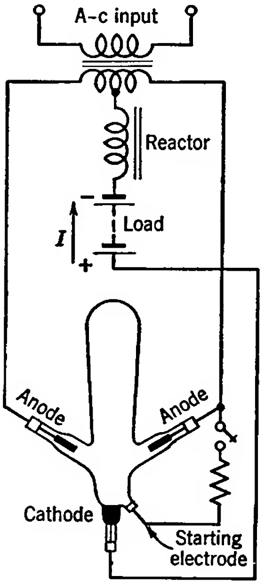


FIG. 6. Simplified circuit of a mercury-arc rectifier.

As the polarity of the input voltage changes, the electron current shifts from one electrode to the other, but, as indicated in Fig. 6, the conventional current always flows up through the battery to be charged. The reactor keeps the current from falling to zero and thus holds the arc.

Sometimes a *single* anode is placed in a mercury-arc rectifier called an **Ignitron.** A starting electrode, called by the trade name **Ignitor,** is employed to "fire" the tube each cycle it is to conduct. This electrode is of some semiconducting material, such as silicon carbide, and one end is permanently immersed in the mercury pool. A small current, when passed from this electrode to the pool, causes the formation of small sparks at the surface of the pool, and these sparks start conduction.

Low-Pressure Gas Thermionic Triodes and Tetrodes. The gas triode contains an oxide-coated thermionic cathode, a control grid, and an anode. Although the grid is called a *control* grid, it can control the operation of the tube only under certain conditions; for other conditions it *loses control.*

In this tube, often called by the trade name **Thyratron,** the control

grid is so designed that for certain anode, or plate, voltages the negative grid is able to block the tube and prevent electron flow from the hot cathode to the plate. If the grid is made less negative or the plate is made more positive than critical values, the grid loses control, electrons suddenly flow to the positive plate, ionization by collision occurs, and the plate passes a current largely controlled by the external resistance in the plate circuit. In the gas tetrode, a second grid is arranged so that the control characteristics can be varied.

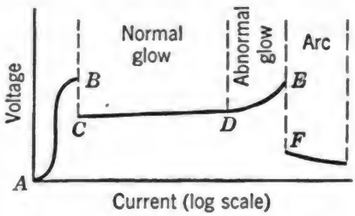


Fig. 7. Phenomena occurring in a two-electrode cold-cathode gas-filled tube.

Two-Electrode Cold-Cathode Tubes. The diodes now under consideration consist of a glass bulb, filled with a stable gas, such as neon or argon at low pressure, and contain two metal electrodes. The small neon bulb is an example of this type.

Fig. 8. A three-electrode cold-cathode gas-filled tube of the general type sometimes used in selective-ringing circuits as explained on page 358. (Courtesy Western Electric Co.)

The characteristics are shown in Fig. 7. Over the region *A–B*, the current that flows is very minute, and no observable discharge occurs within the tube. At *B–C* the voltage required to maintain the current falls to a low value, about 75 volts for a typical tube; also, a slight glow appears as a spot on the cathode. As the current through the tube is permitted to increase over the region *C–D*, the voltage required to maintain the current increases little and the glow gradually increases until at *D* the cathode is covered entirely. Beyond point *D* an increase in current is accompanied by an increase in voltage, and, if the current is increased to *E*, an arc may be formed and the tube will be ruined. This tube is used as a voltage regulator in communication rectifiers.

Three-Electrode Cold-Cathode Tubes.[6, 7] In the tube of Fig. 8 the **main anode** is a wire at the center. A glass tube shields the wire to within a short distance of the end. The **starter anode** and the **cathode** are identical segments of the circular structure shown. They are covered with an oxide coating and in the tube of Fig. 8 are interchangeable.

There are two gaps to be broken down and to be made conducting. The design sometimes is such that the starter gap between the cathode and starter anode breaks down at a lower potential than the main gap from cathode to main anode.

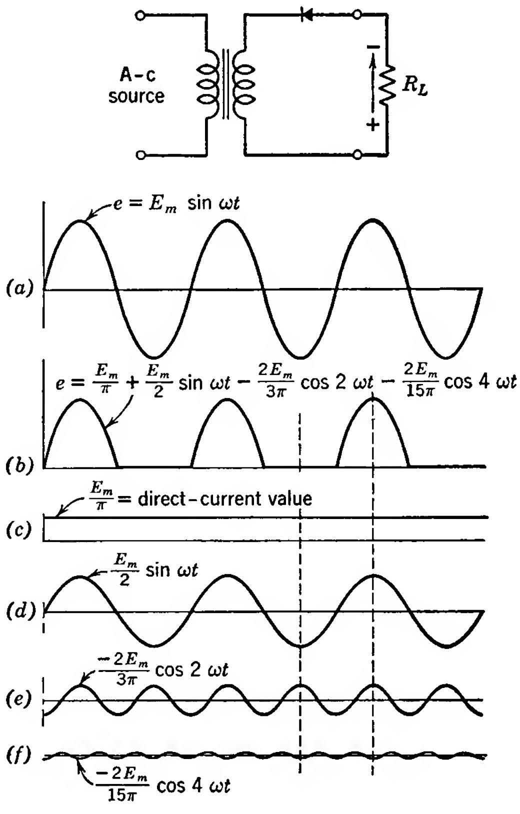


FIG. 9. A half-wave rectifier, and the shapes of the impressed and the rectified waves, *a* and *b*. Waves *c*, *d*, *e*, and *f* are the major components of the rectified wave *b*.

The tube may be used as a relay in this way: Both anodes are made positive but are maintained below the breakdown value and hence conduct negligible current. The device to be operated is placed in series with the main anode. An incoming control or signal impulse is injected in series with the starter anode. This impulse "fires" the starter anode, causing ionization. This supplies sufficient ions to permit the main anode to conduct and operate the device in series with it.

Basic Rectifier Circuits. *Half-Wave Rectifier.* A half-wave rectifier circuit is shown in Fig. 9. Appreciable current flows through the rectifier element only in the direction of the arrow. The sine-wave voltage a, when impressed on the rectifier and load in series, will produce a half-wave current through the load resistor R_L and a half-wave voltage across load resistance R_L, as shown by curve b. If the rectifiers are assumed to be ideal, the equation for this half-wave voltage is

$$e = \frac{E_m}{\pi} + \frac{E_m}{2}\sin \omega t - \frac{2E_m}{3\pi}\cos 2\omega t - \frac{2E_m}{15\pi}\cos 4\omega t \cdots . \tag{1}$$

If the terms e and E_m are changed to i and I_m, then the equation for the rectified current results.

The first term is the desired rectified direct-current component c; the second term is a component d having the same frequency as the source; the third term is a second harmonic e having a frequency twice that of the source; and the fourth term is a fourth harmonic f having a frequency four times that of the source. These separate components are plotted in Fig. 9, and, if they are combined, as by adding at the broken lines, Fig. 9(b) results.

Full-Wave Rectifier. A sine-wave voltage a will produce a full wave such as shown by b of Fig. 10. The equation for the rectified full-wave voltage across the load is

$$e = \frac{2E_m}{\pi} - \frac{4E_m}{3\pi}\cos 2\omega t - \frac{4E_m}{15\pi}\cos 4\omega t - \frac{4E_m}{35\pi}\cos 6\omega t \cdots . \tag{2}$$

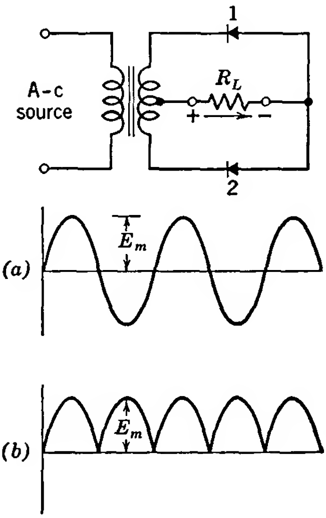


FIG. 10. A full-wave rectifier. The impressed voltage is shown at a, and the rectified current and voltage at b.

This equation can also be used to represent the rectified current. The first term is the desired rectified direct-current component. The other terms are the second, fourth, and sixth harmonics; other harmonics are neglected. These components can be represented in the same manner as in Fig. 9 and then can be combined to give the wave of Fig. 10(b).

Polyphase Rectifiers. When a large amount of power, perhaps a kilowatt or more, is to be rectified, polyphase rectifiers are used.

Three-phase rectifiers are often used with large radio transmitters,[8] and rectifiers of greater numbers of phases are commonly used in large industrial installations.[9]

Bridge Rectifiers. A full-wave bridge rectifier is shown in Fig. 11. For either direction of the applied alternating voltage, current will flow through the resistor as shown. The transformer may be omitted if desired.

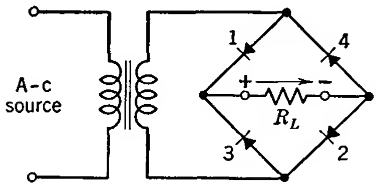


FIG. 11. A full-wave bridge rectifier.

Rectifier Filters. When an alternating voltage is impressed on a rectifier, the output wave is distorted. Distortion is defined[1] as "a change in wave form."

In the usual rectifier unit, the direct component is desired and the alternating components are unwanted. The **rectifier filter** is used to suppress the alternating components.

The battery and the generators of Fig. 12 acting together give the full-wave rectified voltage of equation 2. As an approximation, the internal resistance of the rectifier Z_i can be neglected. Each source may be considered separately, and the magnitude of the current that each source forces through the load R_L can be calculated separately.

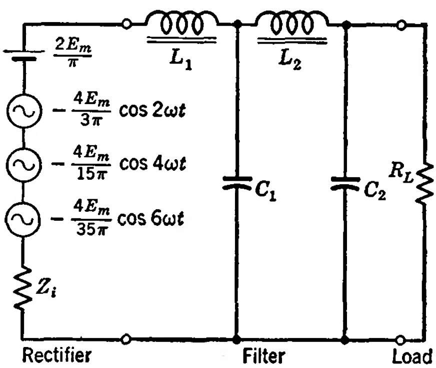


FIG. 12. Equivalent circuit of a full-wave rectifier, filter, and load.

The low-pass filter of Fig. 12 is known as a **choke-input filter.** If the first inductor is omitted, it becomes what is called a **condenser-input filter.** Choke-input filters are assumed in the preceding paragraph. *If the usual condenser-input filter is used, then the rectified (or distorted) voltage wave is not as shown in the preceding figures, and equations 1 and 2 do not apply.* With condenser-input filters, no simple equation is available, because the shape of the rectifier output wave depends on the constants of the filter and on the magnitude of the load resistance.[10] For these reasons, a graphical solution[11] is usually employed for studying rectifier operation with condenser-input filters.

When high-vacuum tubes are used, the filter may have either choke or condenser inputs, but, with large gas tubes, choke-input filters are usually employed. The condenser-input filter draws a high peak current, and this may damage gas tubes.

Cutout in Rectifier Filters. The preceding section implied that, if an inductor were placed next to the rectifier, choke-input filter operation resulted and that, if a capacitor were placed next to the rectifier, condenser-input filter operation resulted. This depends on the magnitude of the inductor and capacitor, and on the resistance of the load.[10] For example, if the load is drawing a *large* direct current, a filter connected as in Fig. 12 may operate as a choke-input filter and the direct voltage will be essentially as given by the first term of equation 2. But, if the same circuit is supplying a *small* value of direct current to the load, the same filter may then operate as a condenser-input filter.

When a rectifier, filter, and load are so related that filter operation is like that with condenser input, the capacitor (even though there may be an inductor between it and the rectifier) draws a high peak current and it charges almost to the peak of the rectified voltage wave; then **cutout** occurs. At this point, the rectifier current falls to zero, and the capacitor maintains the voltage across the filter and load. As a comparison, the voltages across the two types of filters are shown in Fig. 13. It is apparent that the average, or direct, voltage of Fig. 13(a) with choke input is less than in Fig 13(b) for condenser input.

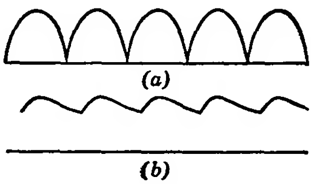


FIG. 13. Rectified waves that result when cutout does not occur (a) and when it does occur (b).

If for any condition of operation, cutout occurs, then the voltage across the load will rise, and the **regulation**[1] will be poor. If certain assumptions are made,[12] it can be shown that for cutout *not* to occur, the relation

$$L_1 = \frac{R_L}{1130} \tag{3}$$

must apply. In this equation, L_1 is the minimum value of inductance in henrys that the first inductor must have, and R_L is the resistance in ohms of the load. It is common practice[13] to use 1000 instead of 1130 in equation 3.

A resistor, called a **bleeder,** is often connected permanently across the filter output (Fig. 14). If this bleeder has the correct value given

by equation 3, it will prevent cutout. Such a bleeder may be provided with taps, and then it serves as a **voltage divider.** A permanently connected bleeder ensures that the capacitors discharge when the rectifier is deenergized, and the bleeder may prevent injuries to operating personnel.

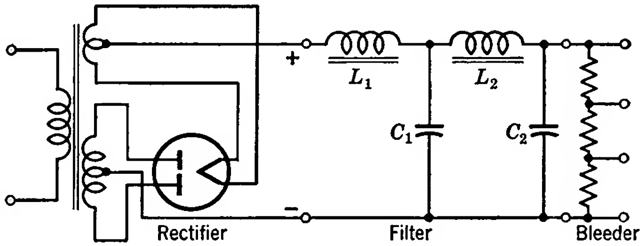


FIG. 14. A typical full-wave rectifier unit. The inductance placed in the positive lead provides better filtering than if placed in the negative lead.

Voltage Stabilizers Using Gas Diodes. Variations in line voltage and changes in the load current will cause the direct voltage output of a rectifier to vary. The cold-cathode gas diode is used to stabilize this output voltage as shown in Fig. 15. The circuit is designed so that the tube is operated over a portion of the characteristic curve of Fig. 7, where the voltage varies but little with changes in current. If the input voltage rises, the cold-cathode gas diode draws a larger current, an increased voltage drop occurs across the series resistor R, and this absorbs the voltage rise. If the input voltage falls, the gas diode draws less current, less voltage drop occurs across resistor R, and the voltage across load resistor R_L remains essentially the same. A stabilizing action also occurs if the load current supplied to R_L varies.

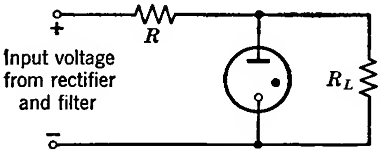


FIG. 15. A cold-cathode gas diode connected to stabilize the voltage impressed across the load R_L.

High-Vacuum Thermionic Three-Electrode Tubes. Triodes were first used as amplifiers in 1906 by De Forest. The first major commercial application was as an amplifier in the **repeaters**[1] of the first transcontinental telephone circuit, opened officially in 1915. Credit is due to Arnold and his Bell System associates for perfecting the triode.[14]

The characteristics of a small triode are shown in Fig. 16. Because

the **grid** of a triode is in a strategic location between the plate and the cathode, the grid has greater influence in controlling the plate current than the plate does. If the plate voltage is e_{b2} and the grid voltage is e_{c2}, the plate current will be i_b. If the plate voltage is reduced to e_{b1}, and the grid voltage is reduced to e_{c1}, the plate current remains i_b. The **amplification factor** is defined[1] as "the ratio of the change in plate voltage to a change in control-grid voltage, under the conditions that the plate current remains unchanged and all other electrode voltages are maintained constant." Thus, the application factor is

$$\mu = \frac{\Delta e_b}{\Delta e_c} = \frac{e_{b2} - e_{b1}}{e_{c2} - e_{c1}}. \tag{4}$$

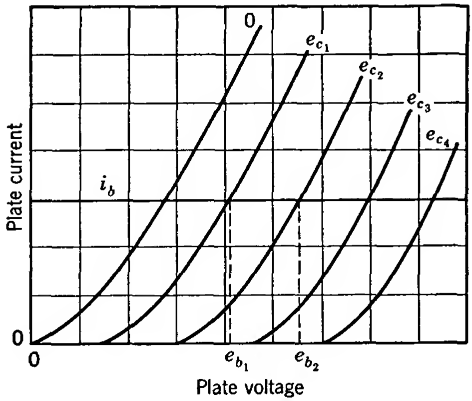


FIG. 16. Characteristics of a small high-vacuum triode.

For triodes the amplification factors may be as low as about 3, and as high as about 100.

When a triode is amplifying a signal, such as speech or music, an alternating current flows from plate to cathode *within* the tube. The opposition to this signal-current flow is the alternating-current **plate resistance**[1] of the tube. The plate resistance r_p of a tube is the reciprocal of the slope of the curves of Fig. 16, and

$$r_p = \frac{\Delta e_b}{\Delta i_b}. \tag{5}$$

For triodes the plate resistance may be as low as about 1000 ohms, and as high as about 100,000 ohms.

When the triode is operated as a non-distorting amplifier, an alternating signal voltage is impressed between the grid and the cathode, and a corresponding signal current flows in the plate circuit. The **control-grid–plate transconductance,**[1] or **mutual conductance,** determines the magnitude of the plate current caused by grid-voltage changes. This factor can be found by dividing equation 4 by equation 5; hence,

$$g_m = \frac{\Delta i_b}{\Delta e_c} = \frac{\mu}{r_p}, \tag{6}$$

and the unit of measure is the mho, or micromho.

High-Vacuum Thermionic Four-Electrode Tubes. Four-electrode tubes, called **tetrodes,** were introduced in this country in about 1928. They are sometimes called **screen-grid tubes,** because they have a second grid, called a **screen grid,** between the control grid and the plate.

The presence of the screen grid shields the control grid from the plate, and the screen-grid tetrode is better than the triode for radio-frequency voltage amplification. With triodes, the amplified signal variations in the plate circuit tend to feed back into the grid circuit and cause oscillations.

In the tetrode, the plate is rendered even less effective (than in the triode) by the presence of the screen grid. For this reason, the amplification factor is about 500 for a typical screen-grid tetrode. The plate resistance also is very high, because the curves of Fig. 16 will be almost horizontal, since a large change in plate voltage will cause almost no change in plate current. A value $r_p = 500{,}000$ ohms is typical.

Although the screen-grid tetrode is for some purposes superior to the triode, the characteristics of the tetrode are impaired by the effect of secondary emission. In amplifiers the screen grid is held at a potential of about 50 to 100 volts positive with respect to the cathode. This voltage and the positive voltage on the plate accelerate the electrons constituting the plate current to such velocities that the electrons cause **secondary emission** (release electrons) when they strike the plate. In an amplifier the plate voltage rises and falls in accordance with the signal being amplified. If the plate potential falls to that of the screen grid, some of the secondary electrons may flow to the screen grid. This phenomenon, in general, is undesired.

High-Vacuum Thermionic Five-Electrode Tubes. In the five-electrode tube, or **pentode,** developed about 1930, a **suppressor grid** is placed between the screen grid and the plate. The suppressor grid is usually directly connected to the cathode. The control grid and the screen grid function as previously explained. The secondary electrons from the plate cannot flow to the screen grid, because to do this they would have to flow through the suppressor grid which is at the potential of the cathode and the cathode is negative with respect to the plate. The so-called suppressor grid does *not* suppress secondary emission but merely forces the secondary electrons back to the plate.

There is even less tendency for feedback than with the screen-grid tetrode, and hence the pentode is a superior radio-frequency voltage amplifier. The additional shielding increases both the amplification

factor and the plate resistance, typical values being $\mu = 1000$ and $r_p = 1{,}500{,}000$ ohms.

Voltage Amplification and Power Amplification. Some amplifiers are designed primarily to increase voltage and are called **voltage amplifiers.** Other amplifiers are designed primarily to furnish power to devices such as loudspeakers and radio antennas and are called **power amplifiers.** A voltage amplifier often increases the feeble signal voltage generated by some device, such as a microphone, until this voltage is large enough to control the output of a power amplifier, which in turn drives a loudspeaker.

Voltage amplifiers use **voltage amplifying tubes** which are small and usually have low direct voltages of several hundred volts or less, on the electrodes. Power amplifiers, however, use **power output tubes** which may be large and may have voltages of thousands of volts on the electrodes.

The so-called **beam-power tube**[15] is used extensively as a power amplifier. No suppressor grid is used in this tube, but, by a special arrangement and because of the shape of the electrodes, the undesired effects of secondary emission are eliminated, and the tube has the characteristics of a pentode.

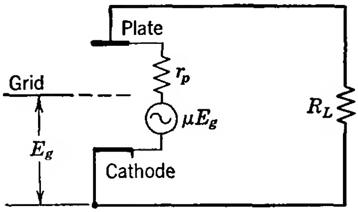


Fig. 17. Equivalent circuit of a vacuum-tube amplifier.

Equivalent Circuit of a Vacuum Tube.[16] When used as a nondistorting amplifier, the signal voltage to be amplified is impressed between the grid and the cathode, and these signal voltage variations cause corresponding signal-current flow in the plate circuit. Because the grid voltage is more effective than the plate voltage in controlling plate-current flow, the equivalent circuit of a vacuum tube as an amplifier is shown in Fig. 17. The alternating signal current that flows through the load resistor is

$$I_p = \frac{\mu E_g}{r_p + R_L}. \tag{7}$$

The signal voltage across R_L is

$$E_L = I_p R_L = \frac{\mu E_g R_L}{r_p + R_L}. \tag{8}$$

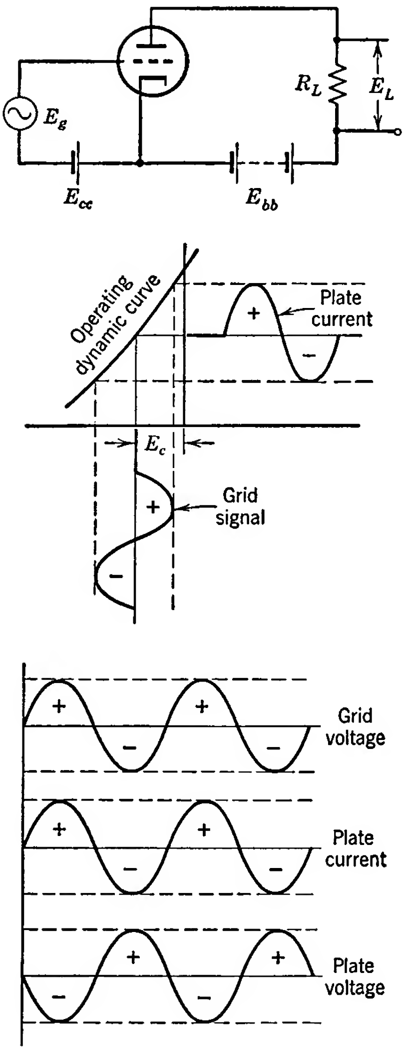


FIG. 18. For an amplifier with resistance load the plate current is in phase with the grid signal voltage, but the plate-to-cathode voltage is 180° out of phase with the grid signal voltage.

The voltage amplification, or voltage gain, of the circuit is

$$A_v = \frac{E_L}{E_g} = \frac{\mu R_L}{r_p + R_L}, \tag{9}$$

which may be expressed in decibels if desired. The power delivered to the load resistor R_L is

$$P = I_p{}^2 R_L = \frac{(\mu E_g)^2 R_L}{(r_p + R_L)^2}. \tag{10}$$

Phase Relations in Amplifiers. The phase relations *for an amplifier working into a pure resistance load* are shown in Fig. 18. The grid of the tube is biased negative an amount E_c by the grid bias source E_{cc}. When the positive half cycle is impressed, the grid is driven *less* negative and *more* plate current flows. This *increase* in plate current causes an increase in the $i_b R_L$ voltage drop in load resistor R_L, and, thus, the voltage from plate to cathode *falls*. When the *negative* half cycle is impressed, the grid is driven *more negative* and *less* plate current flows. This *decrease* in plate current causes a decrease in the $i_b R_L$ voltage drop across R_L, and the voltage from plate to cathode rises. The instantaneous phase relations are as shown in Fig. 18.

Resistance-Coupled Audio-Frequency Voltage Amplifiers. One tube and its associated circuit often give insufficient voltage gain, and hence two or more tubes are connected in tandem, or cascade, to provide additional amplification. A circuit arrangement is shown in Fig. 19. Capacitor C_i is to prevent the flow of direct current from the source through the **grid**

resistor R_g. The electrons returning from the plate through resistor R_k to the cathode cause drops in voltage across **cathode resistors** R_k and **self-bias** the tubes the desired amount. Resistors R_g impress these bias voltages on the control grids of the tubes. **Cathode capacitors** C_k prevent appreciable alternating signal voltage from being produced across R_k by the plate-current signal variations. This prevents feedback (page 300). The **load resistors** R_L are discussed on page

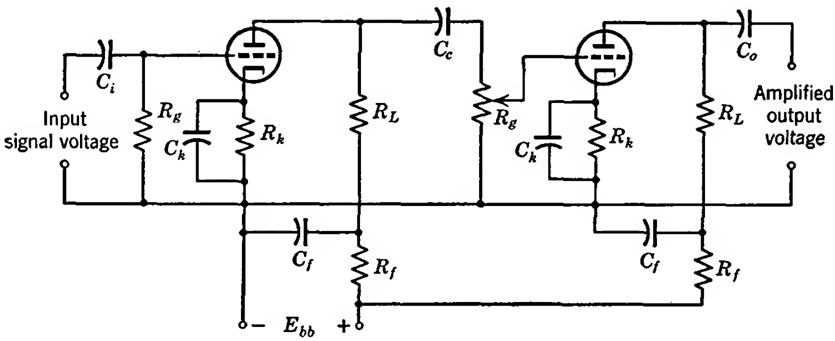


FIG. 19. Circuit of a typical resistance-coupled voltage amplifier using triodes. Typical values are $\mu = 20$, $r_p = 10{,}000$, $R_L = 50{,}000$, $R_k = 1500$, $R_g = 250{,}000$, $R_f = 10{,}000$, all values in ohms; $C_i = 0.05$, $C_k = 25$, $C_f = 10$, $C_c = 0.05$, $C_o = 0.05$, all values in microfarads.

285. Capacitor C_c is the **coupling capacitor** that isolates the plate of one tube from the grid of the next, so that these electrodes can be at different direct potentials. Capacitor C_o is to prevent the flow of direct current into a device connected to the output terminals. Capacitor C_f and resistors R_f are **decoupling capacitors** and **decoupling resistors.** They provide filtering to prevent the flow of alternating signal components from the tubes back into the power supply which is common to both tubes. This prevents **motor-boating,** which is a type of low-frequency oscillation.

Characteristics of Resistance-Coupled Amplifiers. A voltage amplifier must not produce appreciable distortion (page 85). If the tubes are operated correctly, as shown in Fig. 20, the non-linear distortion is negligible. The frequency response (and frequency distortion) is affected greatly by the circuit constants as will now be explained.

The equivalent circuit of one stage of a resistance-coupled amplifier is shown in Fig. 21. The condenser C_g' is the sum of the wiring capacitance C_w and the equivalent grid input capacitance C_g of the

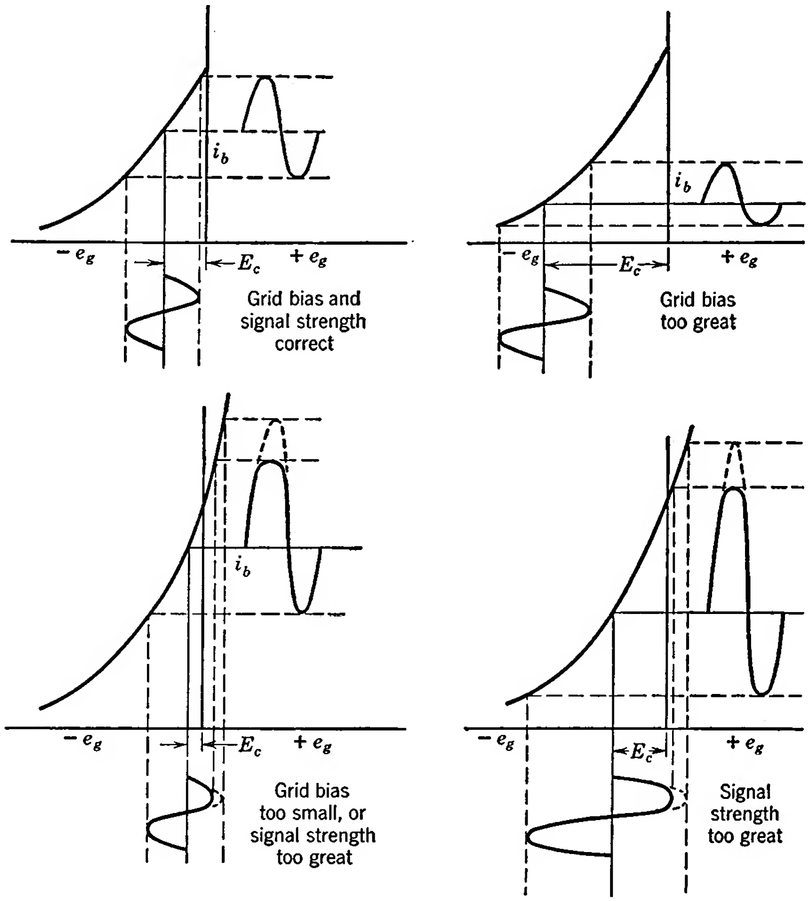


FIG. 20. Either incorrect grid bias or a signal of too great magnitude causes non-linear distortion as indicated.

second tube, given by the equation[12]

$$C_g = C_{gk} + C_{gp}(1 + A_v). \tag{11}$$

In this expression C_{gk} is the grid-to-cathode capacitance and C_{gp} is the grid-to-plate capacitance of the tube, and A_v is the voltage gain per stage. This equation *applies to triodes only*. For tetrodes and pentodes C_g' is the sum of the wiring capacitance and the grid input capacitance which is given in vacuum-tube manuals.

Gain at Intermediate Audio Frequencies. From the values given in Fig. 19 and assuming that C_g' is about 50 micromicrofarads, at an

intermediate frequency of 1000 cycles, the effects of C_c and C_g' are negligible. Then, from equation 9, $A_v = (20 \times 41{,}700)/(10{,}000 + 41{,}700) = 16.1$. The value of R_L used in this equation is the equivalent resistance of R_L and R_g of Fig. 21 in parallel, or $(50{,}000 \times 250{,}000)/(50{,}000 + 250{,}000) = 41{,}700$ ohms. At this frequency the circuit is resistance only, and there is no phase shift except as discussed in connection with Fig. 18.

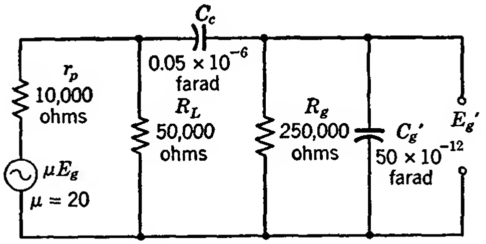


FIG. 21. Equivalent circuit for one stage of Fig. 19.

Gain at Low Audio Frequencies. At low frequencies, such as 50 cycles, the effect of C_g' of Fig. 21 becomes negligible, but C_c must be considered. The frequency at which the gain is 3 decibels below the intermediate audio frequency and the phase shift caused by the circuit is 45° is

$$f_{(\text{low})} = \frac{R_L + r_p}{2\pi C_c(r_pR_g + R_LR_g + r_pR_L)}, \tag{12}$$

where f is in cycles, C_c is in farads, and the resistances are in ohms.

Gain at High Audio Frequencies. At high frequencies, such as 20,000 cycles, the effect of C_c of Fig. 21 becomes negligible but C_g' must be considered. The frequency at which the gain is 3 decibels below the intermediate audio frequency and the phase shift caused by the circuit is 45° is

$$f_{(\text{high})} = \frac{(r_pR_g + R_LR_g + r_pR_L)}{2\pi C_g'R_LR_gr_p}, \tag{13}$$

where f is in cycles, C_g' is in farads, and the resistances are in ohms.

Theory alone predicts that R_L and R_g of Figs. 19 and 21 should be very high so that the alternating current will be low, the internal voltage drop caused by the plate resistance will be low, and most of the available voltage μE_g will appear across the output terminals. Because the direct-current component must flow through R_L, it is

impractical to make R_L more than three to five times r_p, *this rule applying only to triodes.* For triodes, R_g is from about 250,000 to 1,000,000 ohms. For tetrodes and pentodes, the plate resistances are so high that this rule cannot be followed, and typical values of R_L are from 250,000 to 500,000 ohms, with values of R_g about the same as for triodes.

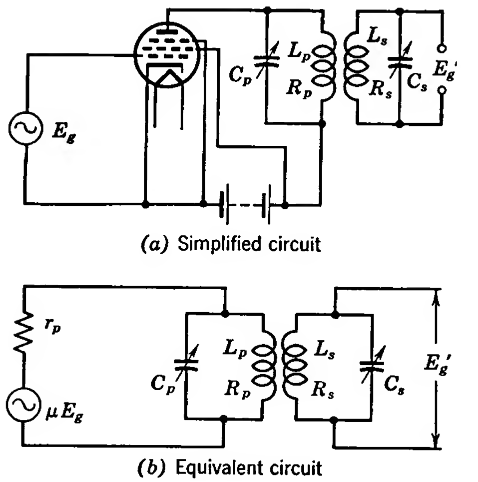


FIG. 22. A radio-frequency transformer-coupled voltage amplifier with tuned primary and secondary.

The equations that have been used were derived[12] using Thévenin's theorem (page 148). Because of the very high internal plate resistance of a pentode (usually over 1,000,000 ohms) and the relatively low load resistance used, equations can be derived using Norton's theorem (page 149). Thus the gain at intermediate frequency is

$$A_v = g_m R_e, \tag{14}$$

where R_e is the equivalent *parallel* resistance of r_p, R_L, and R_g. Often the effect of r_p is neglected because it is so high, and R_e is assumed to equal the parallel resistance of R_L and R_g. This assumption should *not be made with triodes and equation 14 should not be used with triodes.*

Transformer-Coupled Audio-Frequency Voltage Amplifiers. Interstage transformers are sometimes used with triodes. They have a voltage step-up ratio of about 3. If the assumption is made that negligible voltage drop occurs inside the tube, the voltage gain per stage is

$$A_v = \mu \times n, \tag{15}$$

where μ is the amplification factor of the tube and n is the step-up ratio of the transformer.

Radio-Frequency Voltage Amplifiers. Several types of radio-frequency voltage amplifiers are possible; the tuned plate–tuned grid **band-pass amplifier** of Fig. 22 is widely used. The tube operates as a non-distorting amplifier. At the frequency of resonance the impedance in the load circuit is essentially pure resistance. The tube

acts like a generator in series with a plate resistance and a load resistance (Fig. 17).

The equations for the voltage amplification must be derived[17] on the basis of coupled-circuit theory. For pentodes, the voltage gain at the resonant frequency is[18]

$$A_v = \frac{g_m k 2\pi f_r \sqrt{L_1 L_2}}{k^2 + 1/(Q_1 Q_2)}, \tag{16}$$

where g_m is the grid-plate transconductance in mhos, k is the coefficient of coupling, f_r is the resonant frequency in cycles, L_1 and L_2 are the self-inductances in henrys of the primary and the secondary, and Q_1 and Q_2 are the effective energy storage factors of the primary and the secondary. In circuits using pentodes the effective and actual Q are approximately equal.[18]

Classifications of Power Amplifiers. A classification, based on the mode of operation, is as follows:

A Class A Amplifier is[1] "an amplifier in which the grid bias and alternating grid voltages are such that plate current in a specific tube flows at all times."

A Class B Amplifier is[1] "an amplifier in which the grid bias is approximately equal to the cutoff value so that the plate current is approximately zero when no exciting grid voltage is applied, and so that plate current in a specific tube flows for approximately one half of each cycle when an alternating grid voltage is applied."

A Class C Amplifier is[1] "an amplifier in which the grid bias is appreciably greater than the cutoff value so that the plate current in each tube is zero when no alternating grid voltage is applied, and so that plate current flows in a specific tube for appreciably less than one-half of each cycle when an alternating grid voltage is applied."

When the suffix 1 is used it denotes that the grid is not driven positive and that grid current does not flow. The suffix 2 denotes the opposite.

Audio-Frequency Class A1 Single Triode Power Amplifier. The simplified circuit of an audio-frequency power amplifier is shown in Fig. 23. The input impedance of the loudspeaker or other device is transformed into the correct value (page 292) and appears between points 1–2 in series with the tube. Assuming that the load is pure resistance and that the transformer is perfect, then the impedance between points 1–2 will be a pure resistance R_L. The power delivered to the load impedance is given approximately by equation 10.

A convenient method of analysis[19] for a single triode power ampli-

fier is shown in Fig. 24. The direct plate voltage E_b is first selected at some convenient value. The correct grid voltage to use is *approximately*

$$E_c = \frac{0.7E_b}{\mu}. \tag{17}$$

The selected plate voltage and the calculated grid voltage establish the operating point P. The line X–Y is the load line and is drawn

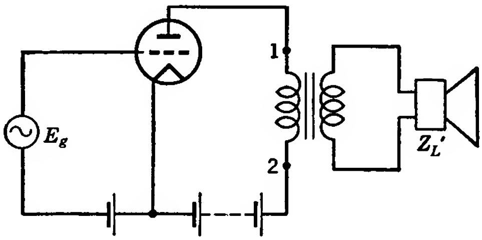


FIG. 23. With an ideal transformer, the load impedance reflected into the plate circuit equals Z_L' multiplied by the turns ratio squared.

through the selected point of operation P at the angle, made with the vertical, $\theta = \tan^{-1} R_L$. It will be noted that in this figure the abscissas are in volts, the ordinates, in milliamperes, and R_L, in ohms. For Fig. 24, the grid bias $E_c = -50$ volts, the plate voltage $E_b = +250$ volts, and the load resistance is $R_L = 3900$ ohms, the grid is driven to zero value, and the plate-voltage and grid-current values are as shown.

For a single tube such as in Fig. 23, the power output is

$$P = \frac{[(I_{\max} - I_{\min})\,(E_{\max} - E_{\min})]}{8}, \tag{18}$$

and the percentage second harmonic distortion is

$$\text{Second harmonic distortion} = \frac{\left[\dfrac{(I_{\max} + I_{\min})}{2} - I_{bo}\right]}{(I_{\max} - I_{\min})} \times 100. \tag{19}$$

The **plate-circuit efficiency,** or ratio of signal power output to direct power input, is low for this amplifier. For maximum power output, the load resistance should equal the plate resistance (page 152). Because of second harmonic distortion, caused by the curvature of the dynamic characteristic curve of a triode, the relation just mentioned cannot be used. Experience has shown that for **maximum "undistorted" power output** the load resistance should be about twice the

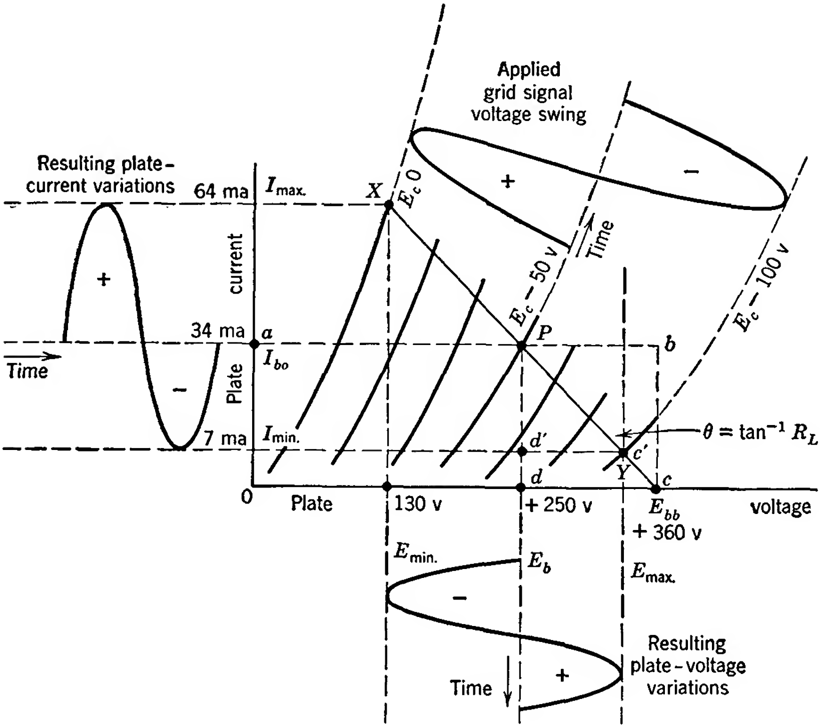


Fig. 24. The applied grid-signal voltage swing causes the plate current and plate voltage to vary as shown. The total power delivered by the plate supply source to the tube and resistance R_L in the plate circuit is represented by the area *oabc*. The power delivered to the tube alone is given by the area *oaPd*. The alternating signal power output of the tube is shown by the area of the triangle *d'Pc'*. If a transformer is used as in Fig. 23, then the amount of direct-current power required is reduced because the tube is loaded with a reflected resistance instead of a resistor R_L.

plate resistance. As used here, "undistorted" means that the nonlinear distortion is less than about 5 per cent.

The **dynamic curve** mentioned is the curve for the tube and its *load resistance.* This curve can be obtained experimentally or can be calculated.[16] It corresponds to the load line of Fig. 24.

Audio-Frequency Class A1 Push-Pull Triode Power Amplifier. In this amplifier two triodes are operated as a **balanced amplifier** (Fig. 25), or in **push-pull.** The second harmonics, which cause most of the distortion when a single triode is used, are largely canceled out in the primary of the push-pull output transformer. Thus, a lower plate load resistance may be used, and the condition for maximum power transfer (page 152) can be more closely approximated. Two

triodes in push-pull will handle more than twice the power handled by a single triode under comparable conditions.

Calculations for two tubes in push-pull are made by first selecting the direct operating voltage E_b and then calculating the direct grid-bias voltage E_c. For class A operation this bias can be between the value for single-tube operation and one-half that required to produce plate-

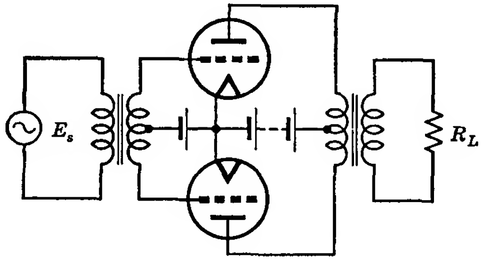


FIG. 25. Two triodes in a push-pull amplifier.

current cutoff when the plate voltage is $1.4E_b$.[19] The maximum power output is then calculated by drawing a vertical line (on a set of static curves such as Fig. 24) at a value $0.6E_b$ until it intersects the zero grid-voltage curve. This establishes the I_{max} value, and then[19]

$$\text{Power output} = \frac{I_{max}E_b}{5}. \tag{20}$$

The correct plate-to-plate load that the output transformer should produce between the two tubes is then four times the resistance represented by the load line drawn as explained. Or, this value is[19]

$$R_{L(pp)} = \frac{(E_b - 0.6E_b)}{I_{max}}, \tag{21}$$

where the units are ohms, volts, and amperes.

Audio-Frequency Class A1 Power Amplifiers Using Pentodes. Tetrodes are seldom used in audio power amplifiers, but pentodes and beam-power tubes are used extensively. The basic principles of operation and the circuits used are similar to those employed with triodes. The methods of calculation are similar but differ in detail.[19]

Audio-Frequency Class B Power Amplifiers. If the definition of a class B amplifier is examined, it is apparent that the output of a tube will be as given by equation 1 and as shown in Fig. 9. Thus, the first alternating term will be the signal to be amplified, and for two push-pull tubes *these will add,* as can be demonstrated by a graphical

analysis.[12] The output wave will contain, however, a series of even harmonic terms, and an analysis will show that *these are suppressed* in the primary of the output transformer. Thus, two tubes in a balanced circuit must be used. In such amplifiers the grids usually are driven positive, grid current flows, and grid-driving power must be supplied.

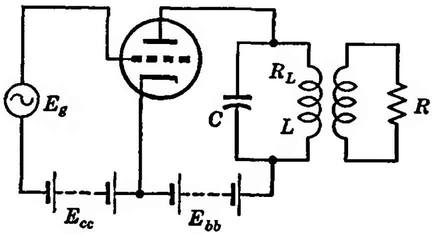


FIG. 26. Simplified circuit for a class B or C radio-frequency power amplifier. The voltage E_g is usually obtained from a tuned circuit. Neutralizing circuits are omitted.

Radio-Frequency Power Amplifiers. The basic circuit of both class B and class C radio-frequency power amplifiers is shown in Fig. 26. The grid is highly biased, often from 2 to 3 times cutoff for class C operation. The grid usually is driven positive so that grid current flows for the **angle of flow** θ_g of Fig. 27. Plate current will flow for the angle of flow θ_p. (Sometimes one-half these angles is taken as the angle of flow.)

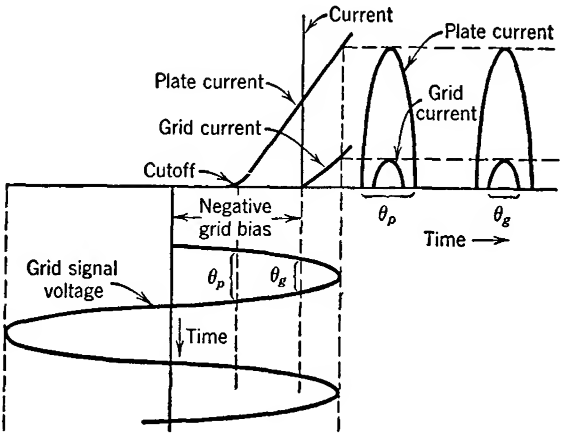


FIG. 27. Relation between grid-signal voltage and currents in a class C radio-frequency power amplifier.

The plate current is badly distorted and is given by a series similar to equation 1. The parallel circuit (sometimes called a **tank circuit**) in series with the plate is tuned to the fundamental signal frequency, and hence this is passed to the output, and the harmonics are suppressed. The phase relations are about as shown in Fig. 18.

Constant-current curves such as those in Fig. 28 are used for calculating the performance of power triodes in class B or class C. The problem is usually to ascertain the operating conditions for maximum

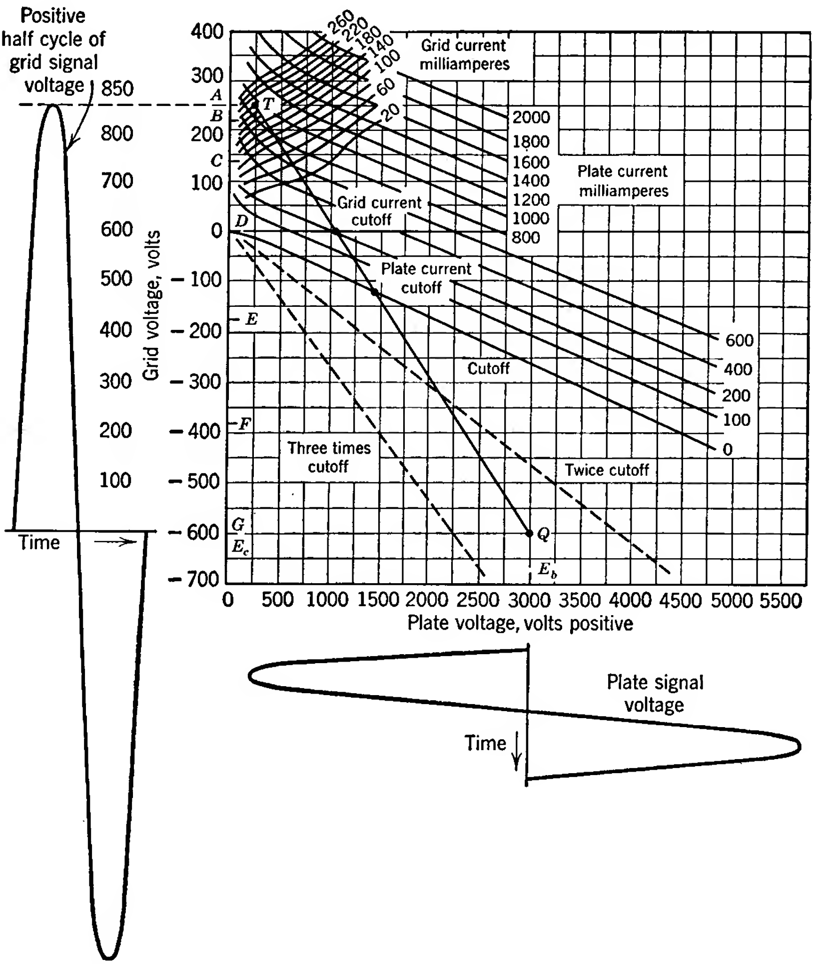


FIG. 28. Constant-current curves and details of operation of class C radio-frequency power amplifier. The peak value of the grid-signal voltage is 850 volts, and the peak value of the plate voltage is 2750 volts.

power output and efficiency and is a cut-and-try process. One method of solution, using a class C amplifier as an illustration, will now be summarized from reference 12.

The quiescent (no signal) **Q point** is selected first. In Fig. 28 this

is at 2.5 times cutoff, and, if the plate voltage is to be +3000 volts, the corresponding grid-bias voltage is −600 volts.

The **tip point** or **T point,** which is the peak positive grid voltage to which the grid is driven and the minimum voltage to which the plate falls, is selected next. In Fig. 28 these values are each +250 volts. This gives a grid-voltage swing of 600 + 250 = 850 volts peak. The plate voltage should not fall below the grid voltage for usual conditions of operation.

When the T and Q points have been arbitrarily selected, the magnitudes of both the grid-voltage signal wave and the plate-voltage signal wave are known. The signal current must be evaluated before conditions of operation can be found. This evaluation is made by a simple method[20] as follows: The grid-driving voltage is divided into equal 15° portions, A, B, C, etc., and the grid-current and plate-current values corresponding to these grid-voltage values are found. Then, the average and the maximum values are given by the equations[20]

$$I_{av} = 0.0833(0.5A + B + C + D + E + F) \tag{22}$$

and

$$I_{max} = 0.0833(A + 1.93B + 1.73C + 1.41D + E + 0.052F). \tag{23}$$

When evaluated from Fig. 28 and calculated as in Table I, the direct plate current is $I_b = 0.181$ ampere and the maximum value is $I_p = 0.331$ ampere. The direct grid current is 0.0226 ampere, and the maximum

TABLE I

EVALUATION OF CURRENT COMPONENTS FOR CLASS C OPERATION

(Method from reference 20)

Angle θ	Cos θ	$E_{g(max)}$ Times Cos θ	Point	Plate Current (Ampere)	I_b from Equation 22	$I_{p(max)}$ from Equation 23	Grid Current (Ampere)	I_c from Equation 22	$I_{g(max)}$ from Equation 23
0	1.0	850	A	0.92	0.46	0.92	0.187	0.094	0.187
15°	0.966	821	B	0.86	0.86	1.66	0.140	0.140	0.270
30°	0.866	736	C	0.62	0.62	1.07	0.037	0.037	0.064
45°	0.707	600	D	0.23	0.23	0.32	0	0	0
60°	0.500	425	E	0	0	0	0	0	0
75°	0.259	220	F	0	0	0	0	0	0
90°	0.0	0	G	0	0	0	0	0	0
				Sums	2.17	3.97		0.271	0.521
				0.0833 × Sums	0.181	0.331		0.0226	0.0434

value is 0.0434 ampere. From these values and assuming that a tuned circuit is a resistance load, the following calculations can be made.

Approximate Grid Dissipation. The alternating grid-driving power input minus the power rectified by the grid is the approximate grid dissipation. The rectified power goes to "charging" the grid-bias battery of Fig. 26 or is dissipated otherwise in the grid-bias source. Thus, the input to the grid circuit is $850 \times 0.707 \times 0.0434 \times 0.707 = 18.5$ watts. The rectified power is $600 \times 0.0226 = 13.6$ watts. The power dissipated in the grid circuit of the tube is $18.5 - 13.6 = 4.9$ watts.

Approximate Plate Dissipation. The direct power input to the plate circuit is $3000 \times 0.181 = 543$ watts. The signal power output is $2750 \times 0.707 \times 0.331 \times 0.707 = 455$ watts. The difference or $543 - 455 = 88$ watts is the power dissipated in the plate circuit.

The Plate-Circuit Efficiency. The plate-circuit efficiency is $455/543 = 0.838$ or 83.8 percent.

Circuit Constants. From experience it is known that the effective Q of the tuned circuit of a class C amplifier, including the effect of the reflected load, should be *about* 12. The equivalent resistance of the parallel load circuit is $R_e = 455/(0.331 \times 0.707)^2 = 8320$ ohms. From equation 16, page 66, and assuming that the Q is 12, the inductance of the tuned circuit can be found, and, knowing this, from equation 13, page 64, the approximate capacitance can be found.

Frequency Multipliers. If the parallel output circuit of a tube in class C is tuned to one of the harmonics, instead of the fundamental, a **frequency multiplier** results; usually these are **frequency doublers** and **frequency triplers.** If the harmonics, and not the fundamental, are to be selected, the tube should be operated so that the desired harmonic is large. An approximate rule is that the plate-current angle of flow θ_p for a frequency multiplier should be

$$\theta_p = \frac{180°}{n}, \tag{24}$$

where n is the ratio of the desired output frequency to the grid-driving frequency. The power output is about $1/n$ that obtained at the fundamental frequency.

Neutralization. Triodes are often used in radio as power amplifiers, oscillators, and frequency multipliers. At these frequencies there is considerable signal fed back from the plate circuit, through the grid-plate interelectrode capacitance, and into the grid circuit; this may cause **positive feedback** and undesired oscillations. **Neutralization** must be used, particularly with triodes, to prevent such oscillations.

Several neutralizing circuits have been developed.[21] One method is to connect an inductor and a capacitor in series between the plate and grid. The capacitor is necessary because the plate and grid are at different direct voltages. This series combination is sufficiently inductive so that it feeds back enough lagging current to neutralize the leading current that flows back through the grid-plate capacitance. Another viewpoint is that the series inductor-capacitor combination in parallel with the interelectrode capacitance causes the impedance between plate and grid to be very high, and hence the net current fed back is very low.

A neutralizing circuit that is highly recommended[21] is shown in Fig. 29. The circuit L–C_1 is adjusted so that it is inductive and a

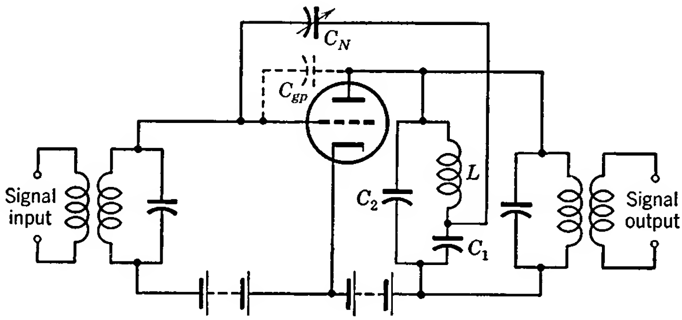


FIG. 29. The neutralizing capacitor C_N feeds back a signal 180° out of phase with that fed back by the interelectrode capacitance C_{gp}.

lagging current flows through C_1. This causes a drop across C_1 that lags the current by 90°. Hence, a voltage exists across C_1 that is 180° out of phase with the plate-to-cathode voltage. When the neutralizing condenser C_N is properly adjusted, as much 180° out-of-phase current will be fed back through it as is fed through C_{gp}, and in this way neutralization is accomplished. Capacitor C_2 is adjusted to parallel resonance with the L–C_1 combination, and hence this parallel combination causes little effect between the points connected.

When tubes are used in push-pull, neutralization is simple. Because the two plate-to-cathode voltages are 180° out of phase, it is only necessary to connect a suitable variable capacitor from the plate of each tube to the grid of the other.

The neutralizing circuits may be adjusted in several ways. One method is to have the circuit in operation, except that the direct plate voltage is not applied. Then the neutralizing circuit is adjusted so

that, with a signal applied to the grid circuit, only a very small signal (as determined with a **wavemeter**) is produced in the plate circuit. The wavemeter is an instrument composed of an adjustable capacitor, an inductor, and a thermogalvanometer in series.

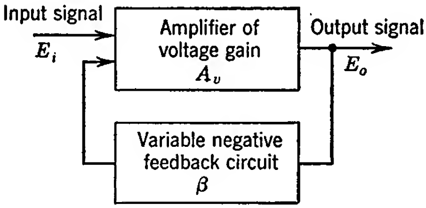


FIG. 30. Schematic circuit of the controlled negative feedback amplifier.

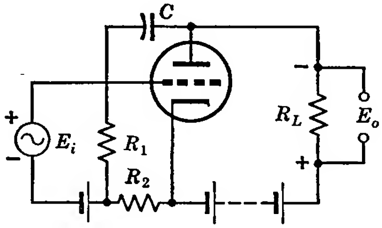


FIG. 31. A circuit for negative voltage feedback.

Feedback.[22, 23] In the preceding section, feedback was undesired, and neutralization was necessary. Controlled feedback is often used, however, with advantage. The basic idea is shown in Fig. 30. Without feedback, the voltage output would be E_iA_v, but with negative feedback the output will be *reduced* to E_o. The controlled negative feedback circuit impresses a voltage βE_o back on the input circuit so that the net input signal is reduced. Thus, $E_o = (E_i - \beta E_o)A_v$, and the

$$\text{Voltage gain with negative feedback} = \frac{A_v}{1 + \beta A_v}, \tag{25}$$

where A_v is the voltage gain without feedback, and β is the **feedback factor.**

Voltage Feedback.[24] Negative voltage feedback is possible with the circuit of Fig. 31. As explained on page 286, the instantaneous voltages will be as marked, with the amplified output voltage across R_L 180° out of phase with the impressed voltage. If capacitor C has low reactance, the current flowing through R_2 will be 180° out of phase with the applied signal voltage, causing negative feedback. The feedback factor is

$$\beta = R_2/(R_1 + R_2). \tag{26}$$

Current Feedback.[24] The circuit of Fig. 32 produces negative current feedback. The positive half cycles of applied grid signal voltage cause the current from plate to cathode (inside the tube) to increase causing a voltage drop across R_C as indicated. The negative half cycles cause the current to decrease, producing a voltage drop across

R_C opposite to that of Fig. 32. This alternating voltage drop across R_C is in opposition to the grid signal voltage and negative feedback results. The feedback factor is

$$\beta = R_C/R_L. \tag{27}$$

Wideband Amplifiers. For applications such as television, amplifiers that will pass very wide frequency bands, such as from 30 to 4,500,000 cycles, are required. If a pentode resistance-coupled amplifier is constructed with low plate load resistances, the high-frequency

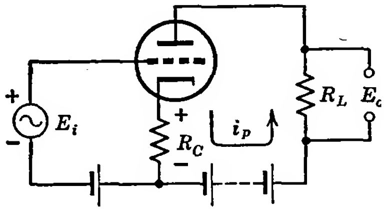


FIG. 32. A circuit for negative current feedback.

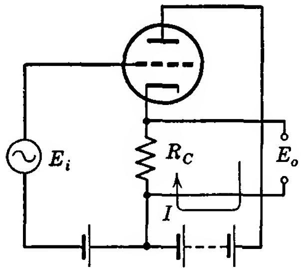


FIG. 33. In the cathode follower, the entire "output" voltage is fed back in series to oppose the impressed signal voltage E_i.

response is extended. **Low-frequency compensation** is achieved by a *parallel R–C* combination *in series* with the load resistor. **High-frequency compensation** is achieved by a small coil in series with the load resistor. These are called **compensated amplifiers.** Other types of wideband amplifiers are possible.[25, 26, 27]

The input capacitance of the following tube plus the wiring capacitance are important factors limiting the high-frequency response (page 289). For this reason, a circuit called the **cathode follower** [28] is often used between two wideband amplifying stages. The circuit arrangement is shown in Fig. 33. The cathode follower has a grid-input capacitance that is much less than the actual capacitance between grid and cathode. The grid-input impedance is correspondingly high. The internal impedance as measured at the output terminals is a low resistance

$$R_{\text{int}} = \frac{R_C}{1 + R_C g_m}. \tag{28}$$

The cathode follower acts like a step-down transformer and can be

used for impedance matching. The output voltage and the input voltage are in phase. The "voltage gain" is less than unity and is

$$A_v = \frac{\mu R_C}{r_p + R_C(1 + \mu)}\cdot \tag{29}$$

A new approach to the design of wideband amplifiers is the so-called **distributed amplifier.**[29] In this amplifier two identical artificial transmission lines, composed of sections of inductance and capacitance

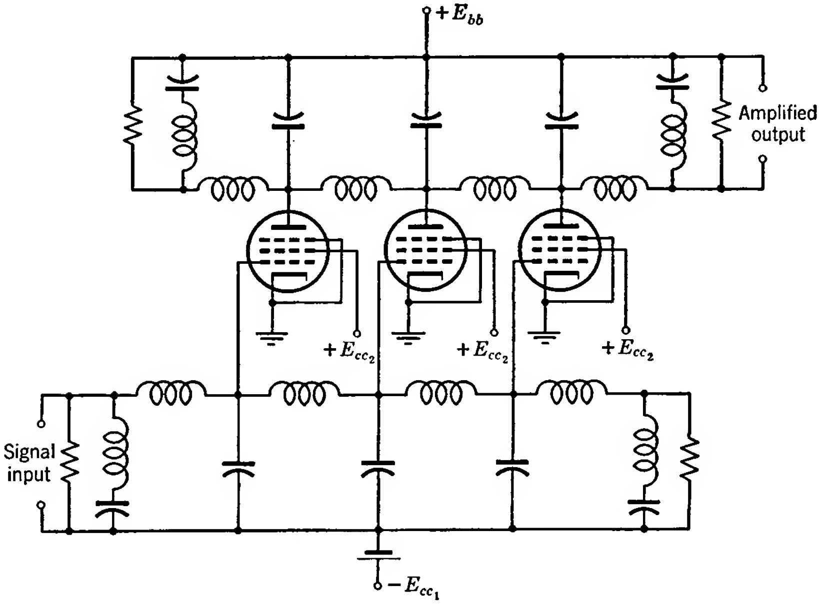


FIG. 34. Simplified circuit of a distributed amplifier that will pass a frequency band of many millions of cycles.

(which may include the interelectrode capacitances), are used. The grids of several tubes are connected in "parallel" across successive sections along one line, and the plates are connected in "parallel" across corresponding sections of the other line, the arrangement being shown in Fig. 34. The lines are, of course, low-pass filters (page 168). They are terminated with m-derived sections (page 182). The cutoff frequency is made considerably higher than the maximum frequency to be amplified, and the input impedance is essentially pure resistance, equaling the iterative impedance of the artificial line. If an input signal voltage is impressed, the first tube amplifies this signal and im-

presses an output signal on the upper line. This signal divides equally, part passing to the left, where it is lost, and part going toward the right. The input signal travels toward the right on the grid line, where the amplifying action is repeated, and the output signal impressed on the plate line again divides, part going to the left, where it is lost, and part going toward the right, where it is joined by the signal traveling to the right from the first tube. In this way, the signals from the successive tubes build up, giving an amplified output. An amplifier, constructed as shown in Fig. 35, was found to have a voltage gain of about 9 decibels from *zero cycles to 40,000,000 cycles, and the band width can be made wider if desired.*

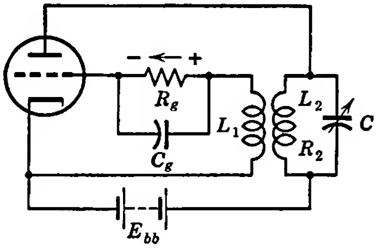


FIG. 35. Basic circuit of a self-biased class C radio-frequency oscillator. Power can be drawn by coupling a coil with L_2.

Inductance-Capacitance (L–C) Oscillators. The frequency of oscillation of Fig. 35 is approximately the frequency of parallel resonance of the L_2–C combination in the plate circuit. The feedback, necessary to sustain oscillations, is from coil L_2 to L_1. The grid is driven positive by an amount determined by the self-biasing R_g–C_g combination and by the coupling between L_1 and L_2. Often the tube is operated in class C. This is a **tuned-plate oscillator.** If capacitor C is placed in parallel with inductor L_1, it becomes a **tuned-grid oscillator.**

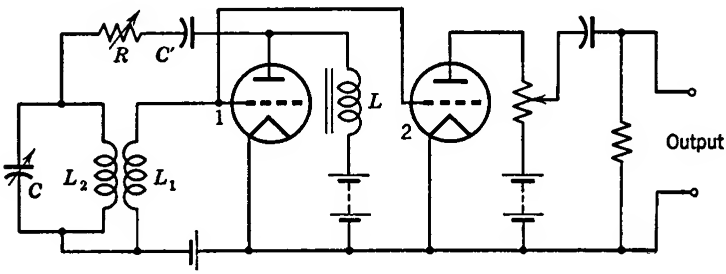
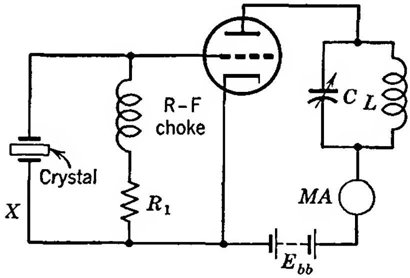


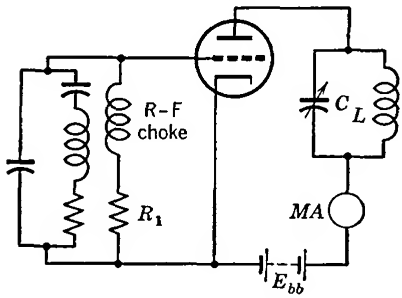
FIG. 36. Simplified circuit of an audio-frequency oscillator.

A convenient oscillator for many purposes is the so-called **resistance-stabilized oscillator** of Fig. 36. Alternating voltage is developed across choke coil L, and alternating current is fed back to the parallel L_2–C circuit through the series R–C' combination. The reactance of C' should be low at the frequency of oscillation, and R should be

reasonably high, often about 50,000 ohms. The second tube of Fig. 36 is an isolating tube known as a **buffer.** This operates as an amplifier but, in addition, prevents changes in the impedance of the connected load from affecting the frequency of oscillation.



(a) Actual circuit



(b) Equivalent circuit

FIG. 37. Actual and equivalent circuits for a triode crystal oscillator. The C–L circuit should be designed to tune to the crystal frequency, a typical value for C being 100 micromicrofarads. The value of R_1 is from 2500 to 25,000 ohms, depending on the grid current that flows. A radio-frequency milliammeter or a radio dial light is often connected in series at point X.

Resistance-Capacitance (R–C) Oscillators. An oscillator may be considered as an amplifier with controlled feedback. Thus, if some of the output from a resistance-coupled amplifier is properly fed back into the input in the proper magnitude and phase, sustained oscillations will result.[30]

Crystal-Controlled Oscillators.[17, 18] Quartz-crystal controlled oscillators are extensively used in radio, a typical circuit being shown in Fig. 37. The quartz crystal is cut and etched so that it is resonant to the desired frequency. When driven at this frequency by signal feedback through the grid-plate electrode capacitance, the crystal oscillates, produces an output voltage between grid and cathode, this voltage is amplified in the tube, and sustained oscillations result. The equivalent circuit of the quartz crystal and its holder are shown in Fig. 37*b*.

Transmission-Line Oscillators.[17, 18] At very high frequencies and ultrahigh frequencies it becomes practicable to use sections of transmission lines to control the frequency of oscillators. Crystals for frequencies higher than *about* 15,000,000 cycles are very fragile and are not in wide use.

The input impedance of a section of a lossless short-circuited transmission line less than one-fourth wavelength long is, theoretically, a value of pure inductive reactance (page 214). The condenser C of Fig. 38 is essentially a short circuit and is adjusted to a point such

that the inductance offered by the line will be in parallel resonance with the capacitance between grid and plate. The inductors shown are radio-frequency chokes to reduce loss of high-frequency energy. Signal from the plate circuit is fed back into the grid-cathode circuit through the interelectrode capacitances; thus, oscillations are sustained.

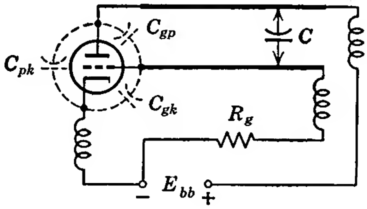


FIG. 38. A simple oscillator using a resonant line to control the frequency. The radio-frequency "chokes" shown reduce high-frequency energy loss.

Cavity Oscillators. A cavity of the general shape to be discussed in this section is, in a sense, composed of many parallel sections of transmission lines that are combined into one metallic structure. At certain frequencies, therefore, the input impedance at the opening of a metallic cavity is pure inductive reactance.

The simplified circuit of a triode oscillator with grounded grid and tuning in both the cathode and plate circuits is shown in Fig. 39*a*. Feedback for sustaining oscillations is through the plate to cathode capacitance. A cavity-type oscillator is shown in Fig. 39*b*. The electronic portion and the cavity portion are in reality part of the same tube structure. The relationship between the two circuits of Figs. 39*a* and 39*b* is evident. The cavities provide the inductances, and the electrodes provide the capacitances, giving an oscillator that will generate frequencies of thousands of megacycles.[31, 32, 33]

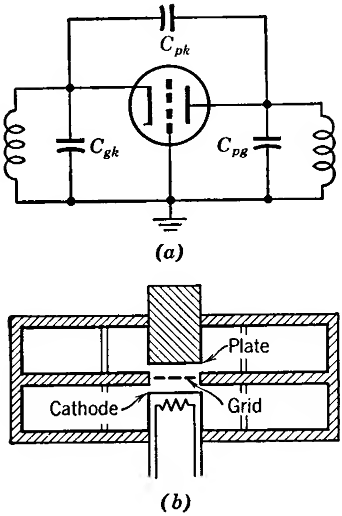


FIG. 39. A simplified oscillator with grounded grid is shown in (*a*). A cavity-type oscillator that is equivalent is shown in (*b*).

A vacuum tube, called a **disk-seal tube**[31] or **lighthouse tube,** is constructed with parallel closely spaced "disk" electrodes such as are shown in Fig. 39*b*. The close spacing reduces the **transit time,** or time required for the electrons to pass from the cathode to the plate. This makes possible high-frequency oscillations. If the transit time is not short and if the frequency is high, an electron

may start out toward the plate, only to arrive there under unfavorable phase conditions. Although the preceding discussion is for the use of disk-seal tubes as oscillators, they are also used at ultrahigh frequencies for amplification and for other purposes, much like ordinary vacuum tubes.[34, 35]

Positive-Grid Oscillators. In certain oscillators for ultrahigh and superhigh frequencies, special triodes are used with the grids *positive* and the plates *negative*. Some of the electrons emitted by the cathode and attracted toward the positive grid do not strike the grid wires but pass between them and travel on toward the plate. The negative plate repels these electrons, and they travel back toward the positive grid, where some of them strike grid wires and some pass between and travel toward the cathode. Here again they are repelled and once again travel toward the positive grid. These oscillations by *some* of the electrons between the electrodes induce voltages of extremely high frequency in the electrodes. The frequency is controlled by the electrode voltages. Such devices are called **Barkhausen oscillators.** When arrangements are made to control the oscillations with short transmission lines they are called **Gill-Morell oscillators.**

Velocity-Modulated Oscillators. The diagram of Fig. 40 is of an early velocity-modulated tube called a **Klystron.** Electrons from the cathode are accelerated by the positive grid and are formed into what may be termed a beam. The electrons that do not strike the grid are further accelerated toward the grid-like structure of the **buncher cavity.** Some electrons are abstracted from the beam by the buncher electrodes, and others pass on to the **catcher,** where again some are abstracted, and the remainder flow on to the collector and back to the cathode. Because a beam of electrons is more or less random in nature, some natural grouping of electrons will occur, for fast electrons will overtake slow ones. These clusters of electrons will induce voltages and currents in the cavities.

This action can be explained as follows: When a group of negative electrons in the electron beam approaches the first grid of the buncher cavity of Fig. 40, these electrons will repel electrons from the first grid and cause them to flow around the cavity walls to the second grid. As the group of electrons in the beam approaches the second grid of the cavity, these electrons drive electrons out of the second grid, around the walls of the cavity, and back to the first grid. This action causes the two grids to have different electric charges on them and results in a potential difference between them. Currents are caused to flow in the cavity walls. If groups of electrons continue to pass rapidly through the grids at regular intervals, high-frequency alter-

nating voltages and currents will be induced in the buncher cavity. Of course, similar action will occur at the grids of the catcher cavity.

The preceding action was based on random bunching of the electrons. If, now, energy is abstracted from the catcher cavity by a small loop placed at the "short-circuited" end where the current is

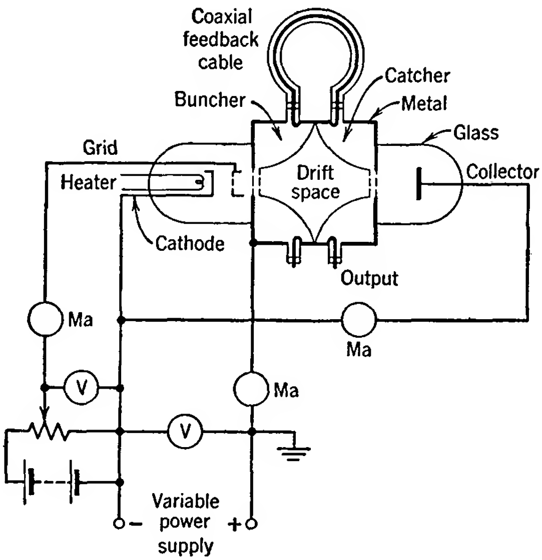


FIG. 40. Simplified drawing of an early Klystron and connections for use as an oscillator. The cavities are marked "Buncher" and "Catcher." Note that the grid is positive with respect to the cathode.

largest (page 211) and if this energy is fed back through the connecting coaxial cable in the proper phase to excite and drive the buncher cavity, an alternating voltage that will intensify the grouping of electrons in the beam will appear between the buncher grids. This additional bunching will cause an increase in catcher voltages and currents, greater feedback will occur, and the oscillations will build up until an equilibrium condition is reached.

The **reflex Klystron** has a *single* resonant cavity, and a **reflector electrode** which is *negative* with respect to the cathode.[36] The electron beam from the cathode is accelerated by the positive grids of the resonator, and some of these electrons flow through the resonator grids and on toward the reflector. The electrons are bunched as they pass through the grids. The negative reflector electrode reflects the elec-

trons back through the resonator grids, and on this return passage energy is abstracted by the resonator from the electrons. By this feedback method oscillations are sustained. The frequency of the oscillations can be varied by changing the potential of either the accelerating voltage or the reflector voltage. This is called **electron tuning.** These tubes readily generate waves a few centimeters long.

The Klystron can be operated as an amplifier. If the signal to be amplified is injected into the buncher cavity through the coaxial cable and small loop, this signal will cause the electron beam to bunch in accordance with it. When these high-speed electrons have reached the catcher cavity they are well bunched and impart signal energy to

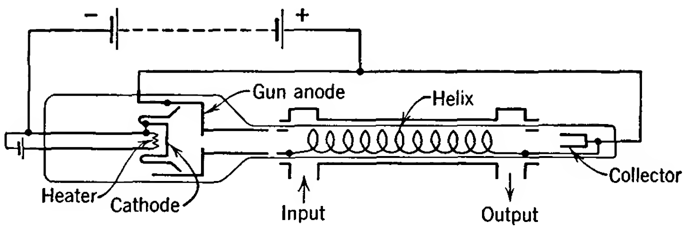


FIG. 41. Diagram of the traveling-wave amplifier tube. (Reference 38.)

the cavity, the energy being abstracted by a loop and coaxial cable. Other Klystron amplifiers are used.[33] Sometimes the cavities are separate from the tube.[37]

Traveling-Wave Amplifier Tube.[38] The electron gun of Fig. 41 shoots an electron beam through a wire helix. The electrons are maintained in a beam by constant magnetic fields produced by the focusing coils. The signal to be amplified is introduced into the traveling-wave tube by a wave guide. This induces a signal voltage in the left end of the helix, and this voltage causes an electromagnetic wave to travel down the helix toward the right end. The velocity of this wave along the *axis of the tube* is much slower than along a straight wire, because the current component must flow around the turns of the helix. The tube is operated so that the velocity of the electron beam is slightly greater than the velocity of the wave along the tube axis. Under these conditions the traveling wave will absorb energy from the electron beam, and signal amplification will result. When the amplified signal wave reaches the right end of the tube, the signal energy is transferred from the helix to the wave-guide output.

The traveling-wave tube has an outstanding characteristic. It will amplify over the very wide frequency range of *800 megacycles,* the bandwidth limits being the usual −3 decibels at the upper and lower frequencies. Furthermore, this tube operates at the ultrahigh and superhigh frequencies of several thousand megacycles. This tube, in the form described or in a modified form with the helix omitted,[39] may prove revolutionary in such applications as point-to-point radio, because of the very wide band that it will pass.

The Magnetron.[40] The magnetron has wide application where large amounts of power at ultrahigh and superhigh frequencies are desired. Peak powers of hundreds of kilowatts for a few microseconds are possible, and for this reason the magnetron is used very extensively in radar systems. At present (1949) the magnetron is not important in commercial communication.

The Cathode-Ray Oscilloscope. The **cathode-ray tube** consists of three basic parts, an **electron gun** that produces a fine beam of rapidly moving electrons, a **deflecting mechanism** by means of which the electron beam can be controlled, and a **fluorescent screen** on which the beam produces a visible spot.

The electron gun consists of a cathode that emits the electrons, one or more grids by which the *number* of electrons in the beam can be controlled, accelerating anodes, that cause the electrons to attain rapid motion, and often a focusing arrangement, although sometimes external coils are used for focusing.

The deflecting mechanism often consists of two pairs of parallel plates between which the electrons flow. One pair is in a horizontal plane, and the other in a vertical plane. By impressing voltages between these plates the motion of the electron beam and the trace on the fluorescent screen can be controlled. Sometimes external magnetic deflecting coils are used.

The fluorescent screen consists of a suitable material deposited on the inside of the tube. The type of material used depends on the nature of the trace desired.

Many of the uses of the cathode-ray oscilloscope are so familiar as to need no review. A use mentioned on pages 172 and 174 is for measuring phase angles, as explained in Fig. 42.

Varistors. Varistors are solid electric semiconductors whose resistance varies greatly with the magnitude of the voltage impressed, the current being conducted, or with temperature. Many types are used in communication, several of which will be mentioned.

Copper Oxide Varistors. The use of copper oxide varistors as rectifiers was discussed early in this chapter. They are widely used in

communication for other purposes, such as modulators and demodulators, which will be considered in Chapter 11.

Silicon Carbide Varistors. The copper oxide varistor just discussed is a **non-symmetrical varistor,** its resistance being very low for one

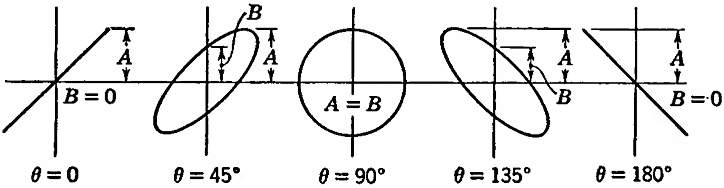


FIG. 42. If the phase relations between the input voltage to a circuit and the output voltage from a circuit are desired, the input voltage is impressed on one set of oscilloscope deflecting plates, and the output voltage is impressed on the other set. For angles between 0 and 180°, sin $\theta = B/A$, the distances B and A being as shown. The figures represent the patterns on the end of the oscilloscope tube.

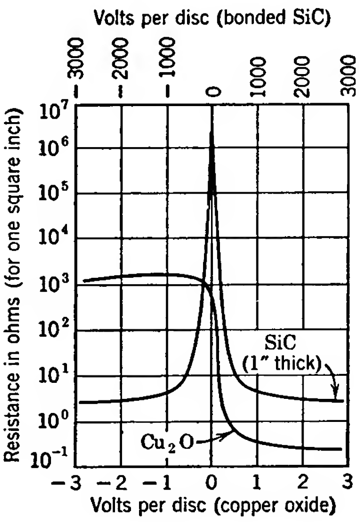


FIG. 43. Comparison of copper oxide and silicon carbide varistors. (Reference 42.)

direction of applied voltage and very high for the other direction as shown by Fig. 43. In contrast, the silicon carbide varistor is a **symmetrical varistor.** For this varistor the resistance varies greatly with the magnitude of the applied voltage but is independent of the direction of the applied voltage.

These varistors consist of granules of silicon carbide firmly embedded in a vitreous ceramic matrix.[41, 42] By a suitable manufacturing process they are formed into strong hard disks, provided with suitable electrodes and vacuum-impregnated with a moisture-resisting material.[41] The change in resistance is not a characteristic of the silicon carbide granules but is caused by the many contacts between adjacent particles.[41]

Thermistors. Thermistors are often classified as varistors, but they require a temperature change to produce the variations in resistance. Thermistors are made of semiconducting materials that have a large variation in resistance with change in temperature.[43]

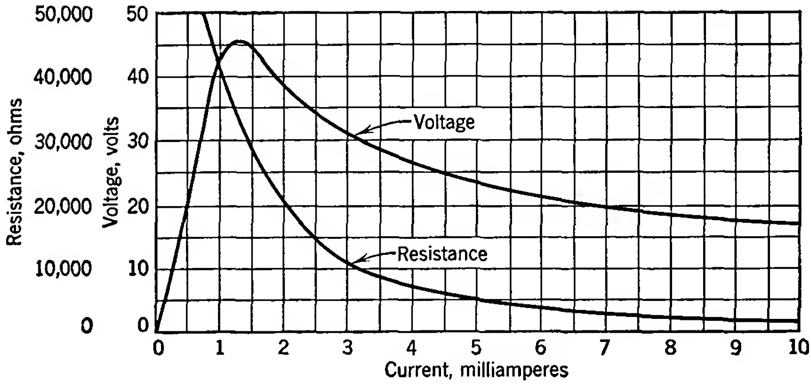
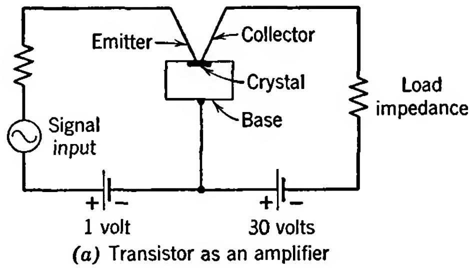


FIG. 44. Characteristics of a directly heated thermistor.

Thermistors are made in many forms. In one type the semiconducting material is controlled by the surrounding temperature. In the *directly heated type,* the current in the circuit in which the thermistor is inserted passes through the semiconducting material and heats it. In the *indirectly heated type* the semiconducting material is heated by a separate heater that is placed in the controlling circuit.



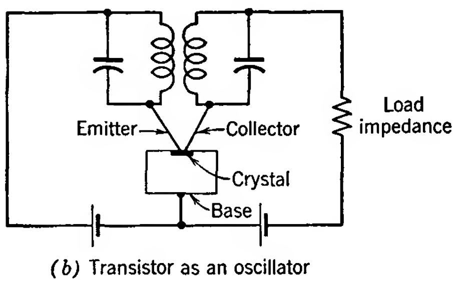


FIG. 45. An amplifier circuit and an oscillator circuit using a Transistor. (Adapted from Reference 45.)

The characteristics of a directly heated thermistor are shown in Fig. 44. As noted, this has a negative resistance characteristic, because the voltage drop decreases with increase in current. Thermistors possess many other interesting characteristics and are used in many ways.[43, 44]

The Transistor.[45] The Transistor shows great promise in communication.[46] Early models consisted of two fine wires whose points press, about 0.002 inch apart, on a germanium crystal.

This will be recognized as a "double crystal detector." The Transistor, however, is used *as an amplifier.*[47, 48] A very feeble input current through one contact can control a much larger current from a source connected to the second set of contacts. Preliminary tests have resulted in amplifications of 100, and currents up to 10 megacycles have been used. This device was developed by Bell Telephone Laboratories. The extent to which the Transistor may replace the vacuum-tube amplifier is not yet generally known (1949). The input impedance of an experimental Transistor is about 1000 ohms, and the output impedance about 10,000 ohms. The power handling capability is 50 milliwatts, and the power consumption about 0.1 watt. Circuits for using the Transistor as an amplifier and as an oscillator are shown in Fig. 45. A coaxial Transistor, with electrodes on opposite sides of a thin germanium element, has been announced.[49]

REFERENCES

1. *American Standard Definitions of Electrical Terms.* 1941. A.I.E.E.
2. Grondahl, L. O. Twenty-five years of Copper—Copper Oxide Rectifiers. *A.I.E.E. Technical Paper* No. 48-66, December, 1947. Also, The Copper Oxide Rectifier. *Elec. Eng.,* November, 1948, Vol. 67, No. 11.
3. History and Application Covered of Copper Oxide Rectifiers. *Elec. Eng.,* April, 1948, Vol. 67, No. 4.
4. Clarke, C. A. Selenium Rectifier Characteristics, Application and Design Factors. *Elec. Commun.,* 1941, Vol. 20, No. 1.
5. Clarke, C. A. The Selenium Rectifier and How It Operates. *Electronic Inds.,* April, 1943, Vol. 2, No. 4.
6. Ingram, S. B. Cold-cathode Gas-filled Tubes as Circuit Elements. *Elec. Eng.,* July, 1939, Vol. 58, No. 7.
7. Bahls, W. E., and C. H. Thomas. A New Gas-filled Triode. *Electronics,* May, 1938.
8. Armstrong, R. W. Polyphase Rectification Special Connection. *Proc. I.R.E.,* January, 1931, Vol. 19, No. 1.
9. Marti, O. K., and T. A. Taylor. Wave Shape of 30- and 60-Phase Rectifier Groups. *Trans. A.I.E.E.,* April, 1940, Vol. 59, No. 4.
10. Stout, M. B. Analysis of Rectifier Filter Circuits. *Elec. Eng.,* September, 1935, Vol. 54, No. 9.
11. Schade, O. H. Analysis of Rectifier Operation. *Proc. I.R.E.,* July, 1943, Vol. 31, No. 7.
12. Albert, A. L. *Fundamental Electronics and Vacuum Tubes.* The Macmillan Co.
13. Dellenbaugh, F. S., Jr., and R. S. Quimby. A series of articles on rectifier design. *QST,* February, March, April, 1932.
14. Ballard, W. R. The High Vacuum Tube Comes before the Supreme Court. *Bell Lab. Record,* July, 1931, Vol. 9, No. 11.
15. Dreyer, J. F. The Beam Power Tube. *Electronics,* April, 1936, Vol. 9, No. 4.
16. E. E. Staff, M.I.T. *Applied Electronics.* John Wiley & Sons.
17. Terman, F. E. *Radio Engineering.* McGraw-Hill Book Co.
18. Terman, F. E. *Radio Engineers' Handbook.* McGraw-Hill Book Co.

19. *R.C.A., Receiving Tube Manual* RC 15.
20. Chaffee, E. L. A Simplified Harmonic Analysis. *Rev. Sci. Instruments,* October, 1936.
21. Doherty, W. H. Neutralization of Radio-frequency Power Amplifiers. *Pick-Ups.* (Western Electric Co.) December, 1939.
22. Black, H. S. Stabilized Feedback Amplifiers. *Elec. Eng.,* January, 1934, Vol. 53.
23. Bode, H. W. Relations Between Attenuation and Phase Shift in Feedback-Amplifier Design. *Bell System Tech. J.,* July, 1940, Vol. 19, No. 3.
24. Schulz, E. H. Comparison of Voltage- and Current-Feedback Amplifiers. *Proc. I.R.E.,* January, 1943, Vol. 31, No. 1.
25. Wheeler, H. A. Wide-band Amplifiers for Television. *Proc. I.R.E.,* July, 1939, Vol. 27, No. 7.
26. Cawein, M. Design of Broad-band Amplifiers. *Electronic Inds.,* September, 1943, Vol. 2, No. 9.
27. Bereskin, A. B. Improved High-frequency Compensation for Wide-band Amplifiers. *Proc. I.R.E.,* October, 1944, Vol. 32, No. 10.
28. Schlesinger, K. Cathode-follower Circuits. *Proc. I.R.E.,* December, 1945, Vol. 33, No. 12.
29. Ginzton, E. L., W. R. Hewlett, J. H. Jasberg, and J. D. Noe. Distributed Amplification. *Proc. I.R.E.,* August, 1948, Vol. 36, No. 8.
30. Terman, F. E., R. R. Buss, W. R. Hewlett, and F. C. Cahill. Some Applications of Negative Feedback with Particular Reference to Laboratory Equipment. *Proc. I.R.E.,* October, 1939, Vol. 27, No. 10.
31. McArthur, E. D. Disk-seal Tubes. *Electronics,* Vol. 18, No. 2, February, 1945.
32. Gurewitsch, A. M., and J. R. Whinnery. Microwave Oscillators Using Disk Seal Tubes. *Proc. I.R.E.,* May, 1947, Vol. 35, No. 4.
33. Spangenberg, K. *Vacuum Tubes.* McGraw-Hill Book Co.
34. Dishal, M. Gain and Noise of Grounded-grid Amplifier at Ultra-high Frequencies. *Proc. I.R.E.,* May, 1944, Vol. 32, No. 4.
35. Jones, M. C. Grounded-grid Radio Frequency Voltage Amplifiers. *Proc. I.R.E.,* July, 1944, Vol. 32, No. 7.
36. Ginzton, E. L., and A. E. Harrison. Reflex-klystron Oscillators. *Proc. I.R.E.,* March, 1946, Vol. 34, No. 3.
37. Bohlke, P. G., and F. C. Breeden. External Cavity Klystron. *Electronics,* July, 1947, Vol. 20, No. 7.
38. Pierce, J. R., and L. M. Field. Traveling-wave Tubes. *Proc. I.R.E.,* February, 1947, Vol. 35, No. 2.
39. Haeff, A. V. The Electron Wave Tube. *Proc. I.R.E.,* January, 1949, Vol. 37, No. 1.
40. Fisk, J. B., H. G. Hagstrum, and P. L. Hartman. The Magnetron as a Generator of Centimeter Waves. *Bell System Tech. J.,* April, 1946, Vol. 25, No. 2.
41. Grisdale, R. O. Silicon Carbide Varistors. *Bell Lab. Record,* October, 1940, Vol. 19, No. 2.
42. Becker, J. A. Varistors: Their Characteristics and Uses. *Bell Lab. Record,* July, 1940, Vol. 18, No. 11.
43. Pearson, G. L. Thermistors: Their Characteristics and Uses. *Bell Lab. Record,* December, 1940, Vol. 19, No. 4.
44. Becker, J. A., C. B. Green, and G. L. Pearson. Properties and Uses of

Thermistors—Thermally Sensitive Resistors. *Elec. Eng.*, November, 1946, Vol. 65, No. 11.
45. The Transistor—A New Amplifier. *Elec. Eng.*, August, 1948, Vol. 67, No. 8.
46. Transistor May Replace Vacuum Tube. *Tele-Tech.*, August, 1948, Vol. 7, No. 8.
47. The Transistor. *Electronics*, September, 1948, Vol. 21, No. 9.
48. BECKER, J. A., and J. N. SHIVE. The Transistor—New Semiconductor Amplifier. *Elec. Eng.*, March, 1949, Vol. 68, No. 3.
49. KOCK, W. E., and R. L. WALLACE. The Coaxial Transistor. *Elec. Eng.*, March, 1949, Vol. 68, No. 3.

REVIEW QUESTIONS

1. What devices are classed as metallic rectifiers?
2. Why is gas used in low-pressure tubes? In high-pressure tubes?
3. Explain the operation of the grid-controlled gas triode.
4. What is the first alternating component in a half-wave rectifier? In a full-wave rectifier? What is the relation of the answers to filtering?
5. What is cutout in rectifier filters? Can it be prevented? How does it affect regulation?
6. Why are tetrodes and pentodes superior voltage amplifiers?
7. Distinguish between voltage amplification and power amplification.
8. Discuss the phase relations in a vacuum tube and its associated amplifying circuit.
9. What determines the low-frequency response of a resistance-coupled amplifier? The high-frequency response?
10. Explain how the decoupling resistors and capacitors of Fig. 19 function.
11. Discuss the application of Thévenin's theorem and Norton's theorem to amplifiers.
12. What are the proper approximate relations of r_p and R_L in triode voltage amplifiers, triode power amplifiers, pentode voltage amplifiers, and pentode power amplifiers?
13. Define class A, class B, and class C operation. What do the suffixes 1 and 2 mean?
14. Define plate-circuit efficiency. What is a typical value for a class A amplifier? For a class C amplifier?
15. What steps are followed in the design of a push-pull class A audio amplifier?
16. Repeat the preceding question for a class C radio amplifier.
17. Explain how frequency multipliers operate.
18. What is the basic principle of neutralization?
19. Describe voltage feedback and current feedback. What is a cathode follower?
20. Describe the so-called distributed amplifier.
21. Explain how a crystal-controlled oscillator operates. Repeat for a transmission-line oscillator.
22. Briefly describe cavity oscillators and enumerate several types.
23. Explain the principle of operation of a positive-grid oscillator. Name a practical application.
24. Explain the principle of operation of a basic Klystron, and of the reflex Klystron.
25. Discuss the traveling-wave tube. What is its outstanding characteristic?

26. Briefly explain how the magnetron operates.
27. What are varistors? Name several types. Give a few applications.
28. Repeat the preceding questions for thermistors.
29. What is a Transistor?
30. Discuss the role the Transistor may have in future telephone systems.

PROBLEMS

1. Calculate and plot the resistance characteristics of the crystal of Fig. 2.
2. Plot the various terms of equation 2 and prove graphically that they combine to give the wave of Fig. 10(*b*).
3. In Fig. 12, the maximum value of the wave $E_m = 475$ volts, and the fundamental frequency is 60 cycles. Capacitors C_1 and C_2 are each 4 microfarads. Inductor L_1 is chosen as 50 per cent greater than the minimum value that will keep cutout from occurring with $R_L = 5000$ ohms. Inductor L_2 is the same as L_1. The resistances of these coils are 150 ohms each. Calculate the approximate magnitude of the voltage that each component causes to exist across load resistance R_L. Justify any assumptions that must be made.
4. Calculate the percentage ripple in Problem 3.
5. Equation 12 gives the low frequency at which the gain of a resistance coupled amplifier is down 3 db, and at which the phase shift is 45°. What will be the frequency for the amplifier of Fig. 19? Does a phase shift of 45° mean that phase distortion is occurring?
6. For the amplifier of Fig. 19 calculate the high frequency at which the amplification is down 3 db. In Fig. 19 triodes were used. How would the amplification compare if two pentodes were used in an appropriate circuit? How would the frequency response compare?
7. Derive equation 18.
8. For the amplifier of Fig. 24, calculate the power output, the plate-circuit efficiency, and the second harmonic distortion.
9. Calculate the angle of flow for both the grid and plate currents for the problem starting on page 296.
10. Repeat the calculations for the class C amplifier starting on page 296, but with a Q point of $E_c = -600$ volts, and $E_b = +2500$ volts.
11. A frequency tripler is to be made using a tube that will put out 100 watts of power in class C. What should be the approximate angle of flow, and approximately what power output will be obtained?
12. Derive equations 26 and 27.
13. Derive equation 29.
14. In the oscillator of Fig. 36, the tubes have amplification factors of 13.8 and plate resistances of 12,000 ohms. The grid bias is -5 volts, and the plate voltage is $+100$ volts. In operation the oscillations build up until the grid of the first tube is driven slightly positive. For a typical circuit, what will be the maximum alternating output voltage?
15. Present a theory for the operation of a Transistor.

CHAPTER 9

TELEGRAPH SYSTEMS

Introduction. For many years telegraph messages were sent by hand-operated keys. Modern telegraph systems, however, now employ the latest developments in circuits, electromechanical apparatus, and electronics.[1] Many interesting telegraph systems have been developed. The following pages will consider certain of these, and will discuss the recent developments that seem destined to revolutionize the industry. The presentation will be confined to systems using open-wire lines and cables. Radio telegraph methods will be discussed in Chapter 13.

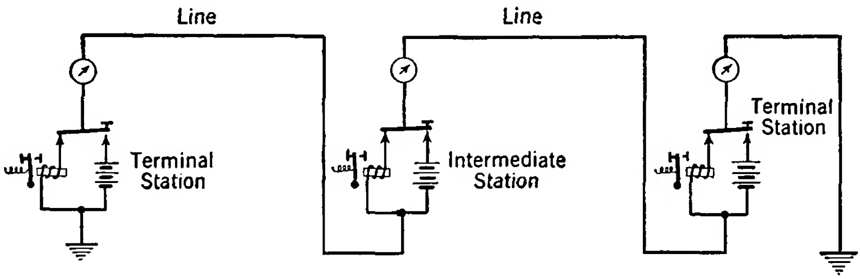


FIG. 1. The open-circuit neutral single-current telegraph system.

Neutral Direct-Current Telegraph Systems. With neutral direct-current systems, signal impulses of either polarity and zero-current spacing intervals are employed.[2] They are often called **single-current systems,** because, in sending a message, current flows in one direction only. Two types of systems have been developed: first, the **open-circuit neutral, or single-current, system** which is used in short "local loops," or circuits; second, the **closed-circuit neutral, or single-current, system** which has had wide use for main-line circuits.

The open-circuit neutral system, the oldest and simplest of all types, is shown in Fig. 1. Current passes over the line only when a message is being sent; no current flows when the line is idle. When the transmitting key at any one of the stations is depressed, the battery at that station is connected to the line and current flows as long as the

key is closed. This current energizes the electromagnets at the other offices, and the sounders respond.

The closed-circuit neutral system is shown in Fig. 2. When a station is not sending, the switch at that station, and all other stations as well, is closed and current flows through the line.

When a station desires to send, the switch at that station (only) is opened and the key is operated manually in accordance with the Morse code. Dots and dashes separated by spaces are sent along the line. Thus, with the closed-circuit system, current flows at all times excepting when spaces are made. If, during the transmission of a message, one operator wishes "to break" the circuit[2] to send a message he can do so by opening his switch. All sounders cease to operate when the

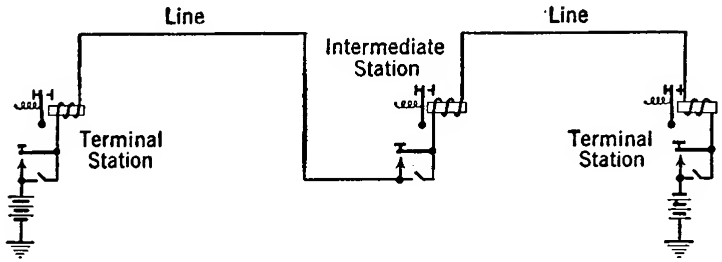


FIG. 2. The closed-circuit neutral single-current telegraph system.

circuit is opened, and the operator, formerly sending, then closes his switch and waits for the message from the station desiring to send.

The open-circuit arrangement is more efficient from an energy standpoint because current flows only when the circuit is closed. However, as seen from Fig. 1, a battery is required at each station, and this cost usually offsets the additional energy cost of the closed-circuit method. In some closed-circuit installations as many as 30 or more intermediate stations have been connected in one line.

Nature of Currents in Neutral, or Single-Current, Systems. As mentioned in the previous section, the currents in a neutral, or single-current, system flow in one direction only and consist of a series of dots, dashes, and spaces.

When the telegraph key is closed, current is allowed to flow through the line operating the relay or sounder to the closed or **marking** position. When the key is opened, the armature of the sounder moves to the open or **spacing** position.

If the line is short and the inductance, capacitance, and leakage are negligible, the circuit may be assumed to consist of resistance only.

The current in such a circuit will be a series of rectangular code impulses. In practice, however, the circuits are long and have large amounts of inductance, capacitance, and leakage; also, the sending and receiving apparatus have considerable inductance and capacitance. These factors distort the signals and may cause them to deviate considerably from rectangular impulses.

A long succession of *uniform* dots and spaces may be considered as being composed of an infinite number of sinusoidal waves superimposed

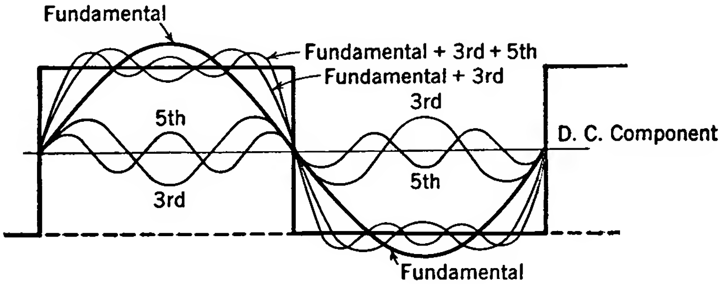


FIG. 3. A periodic rectangular wave, which is similar to a theoretical telegraph signal, is composed of a direct-current component and an infinite number of sinusoidal components. If all components above the third are neglected, the resulting wave is a fair approach to a rectangular wave, at least from the telegraph viewpoint.

on a direct-current component. Pure rectangular waves as in Fig. 3 consist of the fundamental and a complete series of odd harmonics, the equation being

$$i = \frac{4I}{\pi}\left(\sin\theta + \tfrac{1}{3}\sin 3\theta + \tfrac{1}{5}\sin 5\theta + \cdots\right), \tag{1}$$

where I is the maximum value of the rectangular current wave.

With hand operation a Morse operator can send an average of about 30 words, of 5 letters and one space, per minute. This corresponds to a basic frequency of about 12 cycles per second.[3] An expert operator can send 50 words per minute, corresponding to 21 cycles per second.[3] If "square" waves are assumed and if all frequencies up to and including the third harmonic are transmitted, it is evident from Fig. 3 that a Morse circuit can be operated over a channel that will pass from zero to about 100 cycles.[4]

A rectangular telegraph current wave will operate the receiving device in a quick and decisive manner with little distortion, and from this standpoint is desirable. Such a wave, however, would cause considerable **crossfire** (page 331) and **thump** (page 339) in paralleling

communication circuits, because the high-frequency harmonic components would cause greater induction. In practice, series inductance either with or without shunted capacitance is often inserted at the transmitting end to eliminate the higher harmonics in the telegraph

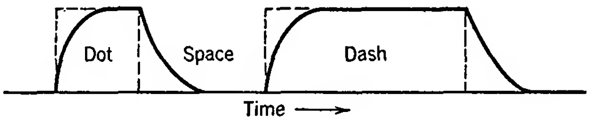


FIG. 4. Because of the inductance and capacitance of telegraph lines and equipment, telegraph signals are rounded off somewhat as indicated.

signals—in other words to "round off" the otherwise rectangular signal waves and produce currents as in Fig. 4.

Neutral, or Single-Current, Telegraph Relays and Repeaters.[5] In the circuit of Fig. 2, the received currents energize the sounders directly. In long lines, however, there is insufficient received energy to operate the sounders. In such instances the line is connected to a

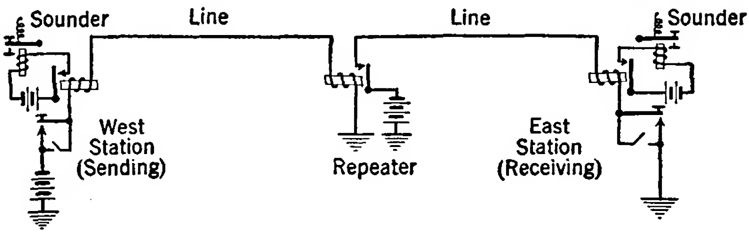


FIG. 5. Early repeater installed near the center of a line to increase telegraphing distances. The feeble signal from line west operates the repeater relay, sending toward the east station a new impulse. Note also that the received signals operate a relay that causes the sounders to be energized from a local battery.

very sensitive relay (Fig. 5) instead of directly to the sounder. The operation of the relay by the incoming telegraph signals will control the battery in the sounder circuit, thus giving satisfactory sounder operation.

A **telegraph repeater** for telegraph transmission over very long distances is indicated in Fig. 5. The attenuated impulses coming in from the sending station are used to operate the relay which sends toward the receiving station an increased amount of power from the battery circuit. This simple arrangement is satisfactory for service in *one direction only;* that is, the receiving station cannot "break" the

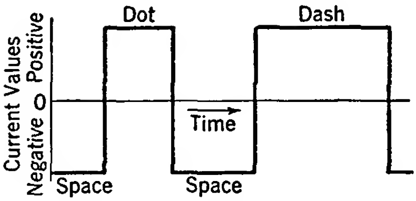


FIG. 6. Shape of signal current in polar, or double-current, system. Inductance and capacitance are assumed negligible.

circuit and send to the other station. A repeater system which provided the breaking feature and permits transmission in both directions over one line wire was early developed.

Polar Direct-Current Telegraph Systems.[2] In a **polar, or double-current, system** the signal impulses are as shown in Fig. 6 if the capacitive and inductive effects are neglected. A dot or a dash is caused by current in one direction, whereas a space is produced by current flow in the *opposite* direction. This accounts for the term "double-current."

The use of double current requires a relay which is *polarized*,[2] that is, which will operate in only one direction for current flow in a specified direction. A relay may be polarized by using a permanent magnet, or by the use of an auxiliary direct-current winding which produces a polarizing magnetic flux. A diagram of a magnetically polarized relay is shown in Fig. 7.

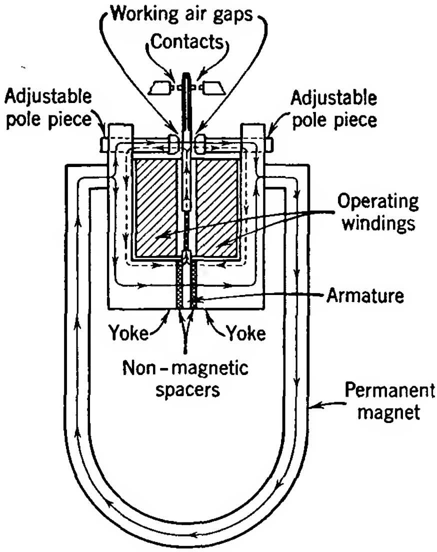


FIG. 7. Diagram of a polarized, or polar, telegraph relay. The magnetic flux produced by the permanent magnet is as shown by solid arrows, and that produced by a given direction of current in the line, or operating, windings is shown by the dotted arrows. The armature, which deflects in one direction or the other in accordance with the telegraph signals in the operating windings, is located at the center. The contacts at the top operate the local circuit to the telegraph office equipment that receives the telegraph signal. (Courtesy Bell System.)

The **polarized,** or **polar, relay** consists of yokes of good magnetic flux-conducting material, such as Permalloy, and one or more operating windings that are connected to the line. A permanent magnet provides the polarizing effect. As indicated in Fig. 7, an armature (also of good flux-conducting material) operating in the air gap carries the contacts for closing the sounder circuit. When a current in one direction

(for instance, corresponding to a space) flows through the circuit, the steady magnetic flux provided by the permanent magnet will be decreased in one yoke and the flux in the other yoke will be strengthened. The armature will then be drawn to the spacing position. When current in the other direction flows, the action is reversed and the armature moves to the marking position, closing the sounder circuit and causing an audible sound corresponding to the signal being sent.

A simplified polar, or double-current, system giving **single operation**[2] or **simplex operation** (that is, transmission in only one direction at a time) is shown in Fig. 8. With the switches as indicated,

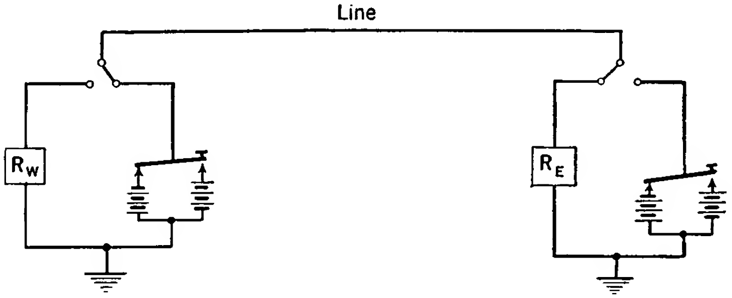


FIG. 8. Double-current, or polar, system for providing single operation. Relays R_W and R_E operate telegraph sounders at the west and east stations, respectively.

transmission is from the west (left) to the east station. The relays R_W and R_E are polarized relays actuating the connected sounders for current in one direction only.

Nature of Currents in Polar, or Double-Current, Systems. The currents in a neutral, or single-current, system and in a polar, or double-current, system are shown for comparison in Fig. 9(*a*). The transmission of current for spacing improves operation as follows: With the neutral, or single-current, system, the closing and opening of the relays and sounders on single current depends on some maximum value M and some minimum value N as in Fig. 9(*a*). If for any reason the current in the circuit is reduced, the curve will be "flattened," M and N will be closer together, the length t_2 of the signal will be decreased, and distortion will result. With the double-current system this does not occur, as illustrated by Fig. 9(*b*). If M_1 and M_2 represent the closing values, and N_1 and N_2 the opening currents for a polarized relay, it is evident that the two closed periods t_1 and t_2 depend but little on the maximum current values indicated. Thus, variations in the line constants with weather changes or variations in current strength due to the addition of other stations or equipment in the line do not cause serious distortion.

A double-current system is less affected by changes in insulation resistance than other systems. The leakage currents passing through the relays (Fig. 2) of the single-current system tend to keep the relays energized at all times, even when sending a space. They must therefore be adjusted as **marginal relays** to operate on *changes* in the magnitude of the current. Thus, as the line leakage changes, a difficult adjustment of the marginal operation of the relays must be made. With the double-current schemes used in practice, the artificial line (page 323) instead of the relays is adjusted to compensate for line changes.

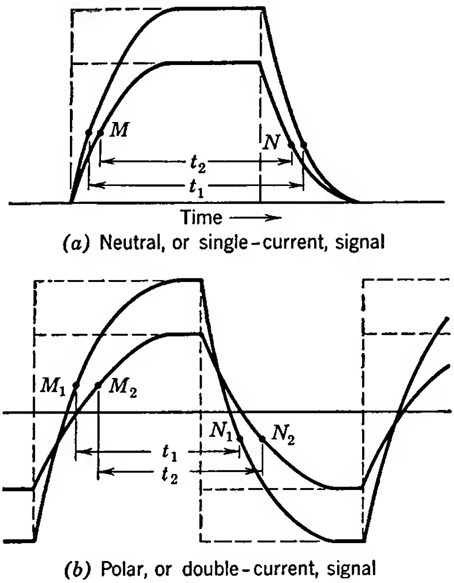


FIG. 9. With the neutral, or single-current, system, a change in the current magnitude has a marked effect on the length of the signal. Thus, in (*a*), the relay would stay closed time t_1 for a large current and t_2 for a small current, and this may cause serious distortion. In (*b*), the current magnitude has little effect on the length of time the relay stays closed.

Duplex Operation.[6, 7, 8] The circuits previously considered provide single, or simplex, operation, that is, a system that can be operated in only one direction at a time. Telegraph systems are in wide use, however, that provide **duplex operation;**[2] that is, simultaneous operation of a channel in opposite directions is possible with **full-duplex operation.**[2] When a duplex system is arranged for **half-duplex operation,** a telegraph channel may be operated in either direction at a time, but not in both directions simultaneously.[2]

The Bridge Duplex System. The bridge duplex system is based on the principle of the Wheatstone bridge, as is evident from Fig. 10. When the operator at the west station presses the key for telegraphing to the east station, the pole-changing relay for sending the "double" current is pulled to the left, sending a current impulse out on the line as shown by the arrows.

If the impedances of coils L_1 and L_2 are identical and if the west artificial line exactly balances the impedance of the connecting line

and the *entire* east set, then I_1 and I_2 will be equal, the bridge will be balanced (page 74), and no current will flow through the west polar relay. Thus, the relay and the sounder at the west sending station will not operate when this station is sending. There are three paths for the current I_1 through the receiving equipment as indicated by the

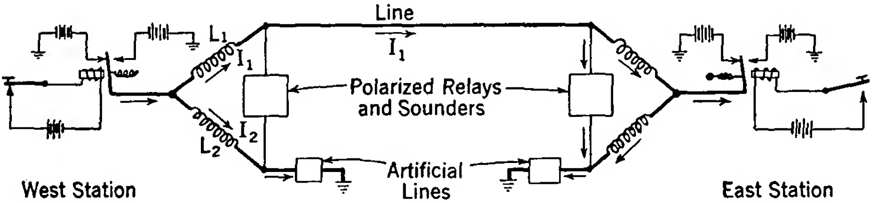


FIG. 10. A simplified bridge duplex system providing simultaneous two-way transmission over one grounded telegraph line wire.

arrows at the east station. Of these three paths, the current through the polarized relay will cause it to operate *if* the current is *of the proper polarity.*

The Differential Duplex System. A simplified differential duplex circuit operating on the double-current principle is shown in Fig. 11.

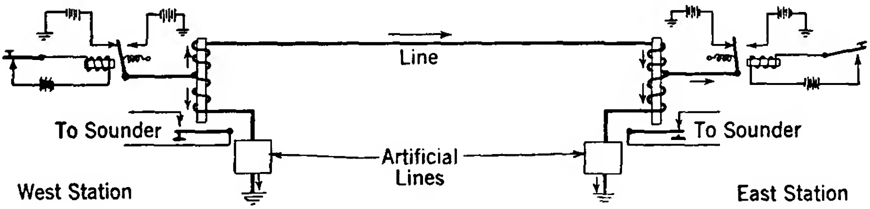


FIG. 11. A simplified differential duplex system providing simultaneous two-way transmission over one grounded telegraph line wire.

The two windings of the differential relays are identical. Thus, if the west station is sending and if the west artificial line exactly balances the impedance of the line and the *entire* east set, the current will divide equally through the two windings. The tendency of the two windings to produce magnetic attraction is then equal and opposite, and the sounder at the sending station is accordingly not operated.

The current flowing over the line to the east station passes through the relay windings as indicated, causing the contacts to close, thus energizing the sounder circuit and producing an audible signal. Either distant sounder is therefore operated by current from the opposite station, and simultaneous two-way transmission is possible over one

telegraph channel, which may be a grounded one-wire line. The differential duplex is widely used and is replacing the bridge duplex type.

Printing Telegraphy. Methods were early devised for transmitting information telegraphically without manual operation and the use of the Morse code. Among the schemes developed were the Creed, Murry, Potts, Siemens, Baudot, Morkrum, Kleinschmidt, and Western Electric.[6, 7, 8, 9, 10] Very little Morse operation is used now, most messages being sent by printing-telegraphy methods.

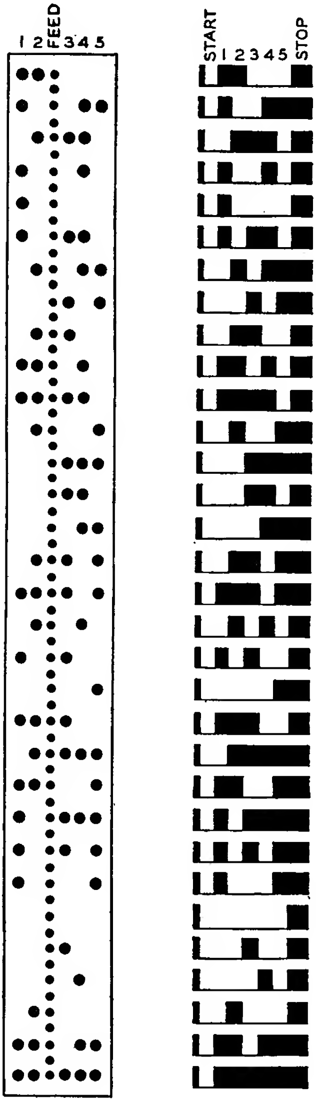
Two systems of printing telegraphy are employed. In **start-stop printing telegraphy** the signal-receiving mechanism is normally at rest and is started in operation at the beginning and stopped at the end of each character transmitted.[2] In **multiplex printing telegraphy**[2] the line circuit is employed to transmit *in turn* characters for each of two or more independent telegraph channels.

Start-Stop Printing Telegraphy. Start-stop printing telegraphy employs an electromechanical device usually called a **teletypewriter** or **teleprinter** at the receiving end of the telegraph channel, and often at the transmitting end as well. Other names, such as **simplex printer, printer,** and **teletype,** are sometimes used. These instruments are complex in mechanical detail, and only the general features will be presented.

Before attempting to explain the start-stop principle, it will be well to consider briefly several possibilities. If a typewriter keyboard were constructed so that when a key is depressed an electrical contact is made, if the printing portion were constructed with an electromagnet operating each printing element, and if one separate line wire were used to connect *each* sending key with each corresponding printing arm, then a message sent at the keyboard would be printed at the distant station. Such a system would, of course, require so many connecting wires as to be of little commercial value.

If a signal code were used, however, the number of wires required could be reduced materially. A number of codes are possible, and the Baudot code shown in Table I has been adopted. Each element of the code is a dot, dashes not being employed. With this code, intelligence could be sent in the following manner: If the letter *A* were to be sent, keys 1 and 2 of Fig. 12 would be depressed, operating electromagnets 1 and 2 at the distant station. Similarly, if an *F* were to be sent, keys 1, 3, and 4 would be depressed, operating the corresponding electromagnets at the distant station. This system would have little commercial value.

Suppose that, instead of the five sending keys being connected to



CHARACTER SENT LOWER CASE	CHARACTER SENT UPPER CASE	SELECTING ELEMENTS OPERATED	CHARACTER RECEIVED LOWER CASE	CHARACTER RECEIVED UPPER CASE
A	–	1 2	A	–
B	?	1 4 5	B	?
C	:	2 3 4	C	:
D	$	1 4	D	$
E	3	1	E	3
F	!	1 3 4	F	!
G	&	2 4 5	G	&
H	£	3 5	H	£
I	8	2 3	I	8
J	'	1 2 4	J	'
K	(	1 2 3 4	K	(
L	)	2 5	L	)
M	.	3 4 5	M	.
N	,	3 4	N	,
O	9	4 5	O	9
P	0	2 3 5	P	0
Q	1	1 2 3 5	Q	1
R	4	2 4	R	4
S	BELL	1 3	S	BELL
T	5	5	T	5
U	7	1 2 3	U	7
V	;	2 3 4 5	V	;
W	2	1 2 5	W	2
X	/	1 3 4 5	X	/
Y	6	1 3 5	Y	6
Z	"	1 5	Z	"
SPACE		3	SPACE	
② CAR. RET.		4	CAR. RET. ②	
③ LINE FEED		2	LINE FEED ③	
FIGURES		1 2 4 5	FIGURES	
LETTERS		1 2 3 4 5	LETTERS	

— NOTE —

① BLOCK SIGNAL INDICATES THAT LOOP CIRCUIT IS CLOSED

② CARRIAGE RETURN OCCURS ON PAGE PRINTERS FOR THIS COMBINATION AND COMMA IS PRINTED ON TAPE PRINTERS

③ LINE FEED OCCURS ON PAGE PRINTERS FOR THIS COMBINATION AND PERIOD IS PRINTED ON TAPE PRINTERS

TABLE I
(Courtesy Bell System.)

the receiving electromagnets by *five* line wires, they are connected to the segments of distributors and the brush arms are connected by *one* line wire. If the arms of Fig. 13 are driven in synchronism and

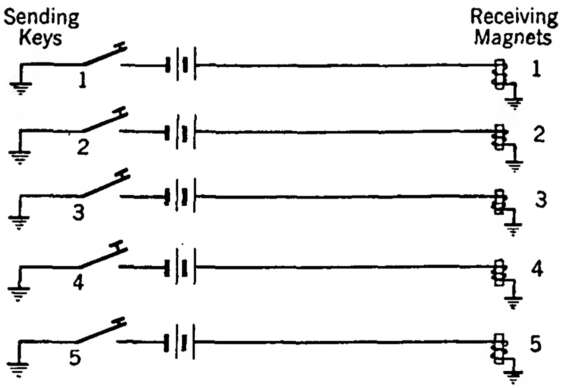


FIG. 12. Telegraph system consisting of five wires used to *illustrate* the principle of the Baudot code and printing telegraphy.

if keys 1 and 2 are closed, electromagnets 1 and 2 will be operated because the brushes pass over *corresponding segments* at the same instant. If these signals are now cleared and keys 1, 2, and 4 are

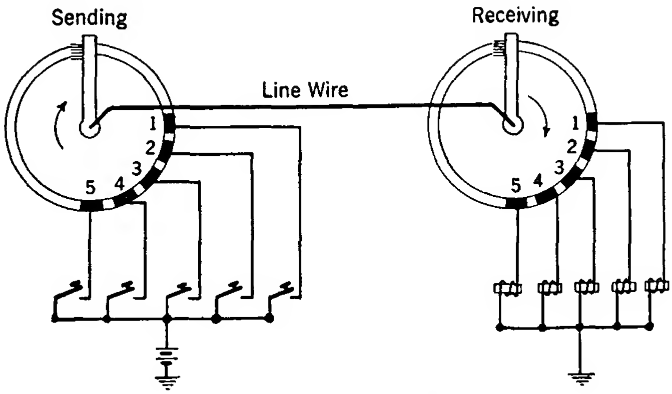


FIG. 13. This figure illustrates how distributors and revolving synchronized brushes eliminate four of the wires of Fig. 12 and make possible printing telegraphy over one line wire.

closed, electromagnets 1, 2, and 4 will be operated on the next revolution. Though more elaborate in detail, both the start-stop and the multiplex systems operate on this principle. Several types of dis-

tributors have been perfected for sending and receiving the impulses; of these the segmented-ring distributors just described have been widely used.

The *elements* of the start-stop equipment are shown in Fig. 14. The sending distributor has connections to a set of six transmitter contacts, and the receiving distributor has connections to a set of seven contacts, five of which are connected to the receiving magnets. The keyboard is similar to a typewriter keyboard.

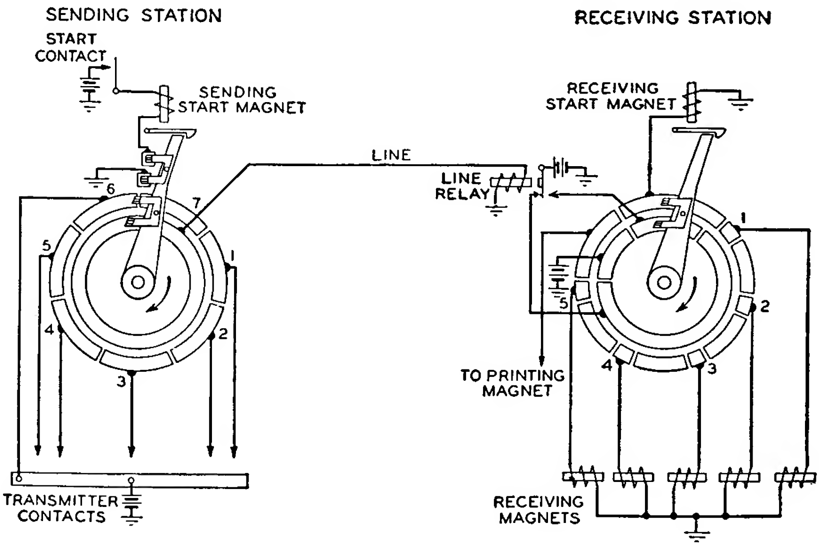


FIG. 14. Simplified circuit of a start-stop system. This figure illustrates the basic *principle* involved. Modern teletypewriters do not operate exactly as indicated.

When a key is pressed, it operates the contact mechanism and energizes the proper segments.[11] Suppose that it is desired to send the letter *A*, which consists (Table I) of the first two units. When the *A* key is depressed segments 1 and 2 will be connected to the transmitter bus by a notched-bar arrangement. At the same time the start magnet is energized, releasing the sending brush arm which is driven by an electric motor. When the arm passes segment 7, a spacing impulse is sent out on the line to the distant receiving teletypewriter, releasing the line relay armature and operating the receiving start magnet which in turn releases the motor-driven receiving brush arm.

The two arms revolve at about the same speed, and, when the sending arm passes over segment 1, which is energized, an impulse is sent over the line to segment 1 of the receiving distributor, which

in turn operates the connected receiving magnet. Similarly, when the sending arm passes over energized segment 2, an impulse is sent receiving magnet 2. The remaining segments 3, 4, and 5 are not energized, and thus no impulses are sent. When the sending brush moves on segment 6 a marking signal is sent, holding the receiving start magnet circuit open and the receiving arm in readiness for the next letter combination. The sending arm is also held for the next letter. The signals received by the magnets operate an arrangement of code selector bars which in turn permit the printing of the letter received.

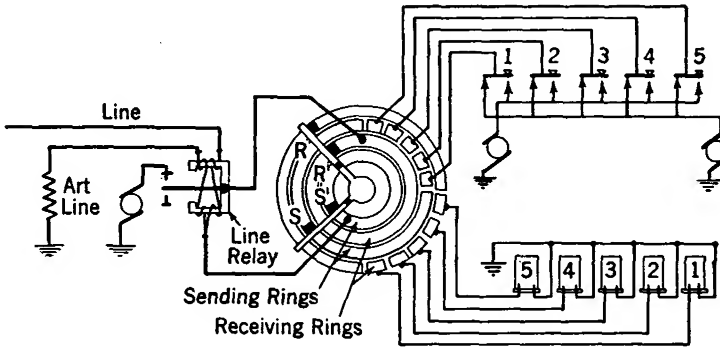


FIG. 15. *Simplified multiplex sending and receiving arrangement.*

Modern teletypewriters[11, 12] operate on this basic principle just explained *but differ* in detail. In particular, a revolving-cam arrangement is used instead of the revolving brush and distributor of Fig. 14. Also, one receiving magnet, instead of five, is used.[13]

Two types of start-stop teletypewriters are available, the **tape printer** and the **page printer.**[11, 12] In the former the received message is printed on a narrow gummed paper strip which can be pasted on the message blanks. In the latter, it is printed on a page, and carbon copies may be made if desired.

Multiplex Operation. With multiplex operation, two *or more* messages are transmitted simultaneously in either or both directions over the same transmission path.[2] Multiplex equipment uses the Baudot code. The sending and receiving brush arms of Fig. 15 are kept in continuous synchronous rotation as long as the sets are in operation. In this respect, they differ from the start-stop apparatus previously described. The sending and receiving elements are combined on one distributor, and two-way operation is provided over a duplexed line.

In this diagram, S and S' and the segments with which these brushes make contact are the sending elements; and R and R' and the corresponding segments are for receiving from the distant station. Since the sending and receiving arms rotate in synchronism, the code combinations must be arranged and the proper segments energized at a constant rate so that when the arm rotates onto these segments they will be ready for sending the correct impulses. A paper tape, on which the message has previously been punched by a keyboard arrangement and fed into the transmitter at a uniform rate, is therefore used to actuate the transmitter equipment. The impulses pass over the line to the receiving magnets as in the start-stop system, and by an arrangement of code selecting bars the signals are translated into the proper letter combination, which is then printed.

With the arrangement just described, two simultaneous messages can be handled over a duplexed grounded line. It will be observed in Fig. 15 that only one-fourth of the face of the distributor plate is in use. It is therefore possible to add three other sets of transmitting segments and three other sets of receiving segments. Then, when the brushes are passing over one set of segments, the signal impulses are being arranged for the other three sets of segments, and with a duplexed line *eight messages are handled in a sense "simultaneously" over one telegraph line.*[14]

Automatic Sending. Using start-stop equipment, the operator may "type" the message directly "onto the line" or the operator may record the message as perforations in a paper tape (Table I). This perforated tape, when fed into an automatic transmitter, sends the telegraph message to the distant station. Several possibilities exist: first, the tape may feed directly from the tape perforating equipment to the automatic transmitting equipment; and second, operators may perforate tape at their convenience, and then the recorded messages may be transmitted continuously.

Telegraph Lines and Cables. Telegraph messages are transmitted over **telegraph channels,** defined[2] as a "path which is suitable for the transmission of telegraph signals between two telegraph stations." These channels are provided over open-wire lines, cables, and by radio (Chapter 13). Telegraph facilities within the United States are provided by The Western Union Telegraph Company and by the Bell System.

The Western Union circuits are designed primarily for telegraph service. The open-wire mileage exceeds 2 million miles, of which about 72 per cent is copper, mostly No. 9 A.W.G., and 28 per cent is iron or steel with No. 8 B.W.G. predominating.[15] Cables used pri-

marily for telegraph service are of the twisted-pair paper-insulated type. For intercity trunk-line service, No. 16 A.W.G., or larger, copper conductors are used. For local distribution purposes the cable conductors are often of No. 19 A.W.G. copper wire.

The Bell System lines and cables are designed primarily for telephone service, but simultaneous telephone and telegraph transmission is possible over many of the circuits.

Telegraph Transmission Theory. In telegraphy, the information transmitted is conveyed by the timing of transitions from one steady-state condition to another.[16] In the single-current, or neutral, system, a dot is recognized as a certain time interval, that is, the difference between the instant of a change from no current to a given current, and the instant of a change from the given current back to no current. In the double-current, or polar, system, a dot is transmitted as a time interval between the instant of transition, or change, from current in one direction to current in the other direction, and the instant of change back to the original condition. Thus in telegraphy the instant of transition is important, and *signal spacings* must be maintained.

The shapes of telegraph signals were shown in Figs. 3, 4, 6, and 9. An analysis of the transmission problems involved can be made on either a transient basis, or on a steady-state sine-wave basis in which the Fourier method of analysis is used.[16] This method was indicated in the consideration of Fig. 3. The transient basis was used for many years and is still used, but the sinusoidal steady-state method[17] has proved to be more practicable. When the sine-wave method is used, the theory of the transmission of telegraph signals is essentially the same as the transmission theory considered in preceding chapters.

The sending speed with Morse operation was discussed on page 318. With teletypewriter operation the standard operating speeds are 40, 60, and 75 words per minute, although faster operation is used to a limited extent. **Telegraph transmission speed** is defined[2] as "the rate at which signals are transmitted, and may be measured by the equivalent number of **dot-cycles** per second, or by the average number of letters or words transmitted and received per minute." A given number of dot cycles per second, each of which consists of an on-off, or a mark-space, signal, is equivalent to twice that number of **bauds**,[2] a baud being one pulse per second. A speed of 60 words per minute requires 360 operations per minute and gives a dot frequency of about 25 cycles per second. If at least the third harmonic is transmitted to produce an approximately square wave (Fig. 3), then the theoretical band width required is about 75 cycles. In practice, band widths of

about 100 cycles are provided for telegraph service. It is emphasized that the figures just given, and those on page 318 are approximate.

Telegraph Distortion. In general, distortion is defined as a change in wave form.[2] **Telegraph signal distortion** is defined[2] as "the deviation of signals from a facsimile reproduction of the impressed signals as regards the time of beginning and the time of ending of the corresponding individual components, aside from the average lag of the signals."

In most manual or Morse telegraph systems the incoming signals operate a relay which in turn closes a local circuit (Fig. 5) actuating the sounder. In teletypewriter systems a local circuit energizes the receiving magnet. The volume of the received telegraph signal is *relatively* unimportant; the sound produced or the letter printed will be just as satisfactory (within limits of good operation) with weak as with strong received signals, provided that the lengths of the signals are not distorted. The amount of distortion can be determined by measuring the lengths of the impulses.[18, 19, 20]

Telegraph distortion is of two types,[21] **systematic** and **fortuitous.** Systematic distortion may be divided into **bias** and **characteristic** distortion.

Bias. This consists of a lengthening of the marking impulses and a corresponding shortening of the spaces, or vice versa. If the signal impulses are lengthened, it is called **positive bias;** if the spaces are lengthened and the signals shortened, it is **negative bias.** This form of distortion is usually due to differences in battery potentials, improper relay adjustment, steady duplex unbalance currents, earth potential differences, and leakage currents.

Characteristic Distortion. This is distortion peculiar to a given signal combination resulting from the *electrical characteristics* of the particular line and relay being used. It is called characteristic distortion if it changes in neither sign nor magnitude when the functions of the spacing and marking signals are interchanged.

Fortuitous Distortion. This is an erratic (as contrasted with a systematic) lengthening and shortening of the impulses or spaces due either to chattering or sparking of relays or to interfering currents from external sources.

There are three general sources of external electrical interference. They are from paralleling telegraph circuits, called **crossfire;**[2] from paralleling power circuits; and from natural sources. Crossfire transfers telegraph signals from one circuit to another through the electric and magnetic fields. It can be controlled by transpositions in metallic circuits but is particularly bothersome in grounded circuits.[22]

The aurora borealis often is accompanied by one of the worst sources of natural disturbances.[23] During such displays (as well as at other times[24]) very large earth currents flow, and these produce high differences of potential between the ends of grounded telegraph circuits. These earth currents and resulting potentials are sometimes so severe as to prevent service, operate protective devices, and damage cables and equipment. Such disturbances often affect large sections of the country simultaneously.

Lightning also is a cause of interference with telegraph operation and by either direct strokes or induced potentials may affect transmission or damage equipment. Trouble is also caused by storms driving charged particles of snow against the line, thus causing the line to assume a potential and currents to flow. Similar interference is caused by high winds driving charged sand and dust particles against the line wires.

Regenerative Repeaters. Repeaters for neutral, or single-current, telegraph systems were discussed early in this chapter. Repeaters are also used on circuits operated by printing telegraph equipment. In addition to increasing the signal strength, the repeaters reshape the signal to compensate for distortion, and longer transmission distances are possible. This reshaping is accomplished by **regenerative repeaters,** defined[2] as "a telegraph repeater which receives mechanically sent, electrically transmitted signals and resends them in substantially perfect form."

Where circuits are composed of a number of repeater sections, the distortion of each section is cumulative, and a point is reached where distortion alone limits the transmitting distance. The regenerative repeater eliminates distortion and strengthens the line current by using a small part of the middle portion of the received impulse to determine the nature of the outgoing regenerated signal. The relation between the distorted received signal and the corrected impulses in one system is indicated in Fig. 16.

Regenerative telegraph repeaters are of several types, including repeaters operating on the non-rotary, rotary, and electronic principles.[15, 25, 26, 27, 28, 29, 30]

The Varioplex System.[15, 31, 32] Varioplex equipment, first installed by Western Union about 1940, causes a multiplex system or other suitable telegraph transmission facility *to be fully loaded* at all times. The varioplex has been extensively used with multiplex systems, and a multiplex channel will be used to explain the basic principle of operation of the varioplex.

With a multiplex system, four two-way telegraph channels are made

available over one telegraph circuit which may be a single grounded wire of an open-wire line. The transmission is not simultaneous; in fact, use of the wire is shared on an equal time basis by the four two-way channels. If one channel becomes idle for any reason, the time assigned to it is lost, and hence the transmission circuit, or other facility, is not worked to maximum efficiency.

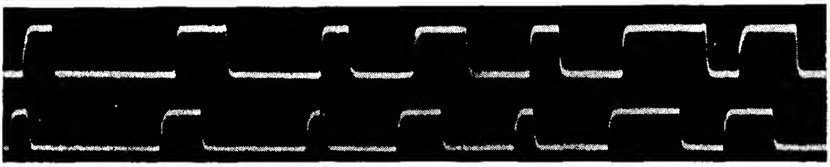


Fig. 16. Oscillograms of distorted telegraph signals (below) and the regenerated signal (above). (Reference 26.)

With the varioplex arrangement a number of subchannels are connected through the varioplex equipment to a multiplex system (or to a carrier telegraph system). Each subchannel terminates in printing telegraph equipment. As many as 36 subchannels have been connected through varioplex equipment to one four-channel multiplex system that operates over one grounded transmission line wire.

An operator of a subchannel sends at will. The incoming message signals operate a tape reperforator in the varioplex equipment and temporarily store the message. The varioplex apparatus connects the various subchannels in sequence to the multiplex system, giving each subchannel an opportunity to send a character. However, a "sensing circuit" causes the transmitting sequence equipment to "skip" any subchannel that at the instant has nothing to send, and to transmit immediately from a channel that does have a character to send. In this way, the transmitting speed is increased or decreased to accommodate the number of subchannels that are in operation, and the transmission circuits are loaded to maximum capacity a large percentage of the time.

Varioplex has found wide application in providing printing **telemetered telegraph service** to customers who have considerable traffic to send, but who could not use a circuit continuously. The subchannel provided to such customers is available to them instantaneously and at all times, and they pay on the basis of the words sent as measured by a telegraph character counter.

Carrier Telegraph Systems. Carrier systems are used by both the Bell System (page 411) and Western Union to increase the message

handling capacity of transmission circuits. Two basic methods are used, **amplitude-modulated carrier systems** and **frequency-modulated carrier systems.** Because both modulation and carrier systems will be considered in detail in Chapter 11, the present treatment of carrier telegraph systems will be condensed.

Amplitude-Modulated Carrier Telegraph Systems.[15, 16, 33–37] A different carrier frequency, higher than the usual telegraph frequencies, is assigned to each incoming telegraph channel from printing telegraph equipment. In effect, carrier-frequency energy is caused to flow on the line to the distant receiving station in accordance with the

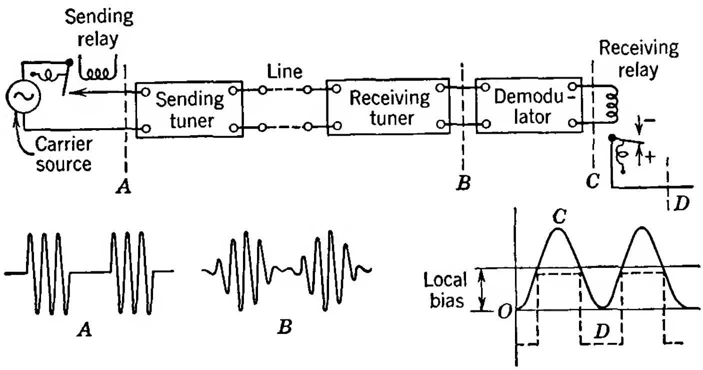


FIG. 17. Basic diagram of one channel of an amplitude-modulation carrier telegraph system. The "line" in the upper diagram refers to the telegraph transmission line between the transmitting and the receiving stations. Many channels, such as the one shown here, each operating at a different frequency, will be connected in parallel to this transmission line. In the lower figures are shown the approximate shapes of the currents at the corresponding points above. (Reference 37.)

incoming telegraph signals. Many such channels, each employing a carrier of different frequency, are connected in parallel to the telegraph line which usually is a two-wire, or metallic, circuit that is designed to transmit the carrier-telegraph signals. Because the amplitude of the carrier wave is controlled or **modulated** by the incoming telegraph signals from the teletypewriters, the term **amplitude modulation** is used. Sometimes this is termed **on-off modulation.** At the receiving station the high-frequency signals are demodulated by suitable rectifying equipment, and the resulting current impulses are used to operate the connected printing equipment.

The details of the several carrier telegraph systems vary, but the basic idea is shown in Fig. 17 which is for *one carrier* channel. The incoming impulses from the sending telegraph equipment cause the

sending relay to connect the carrier source and transmit carrier impulses, such as in *A*, into the sending tuner. This "smooths out" the impulse by removing undesired high-frequency harmonics (see also page 318) and impresses an impulse such as *B* on the line, together with other such impulses from other channels. The receiving tuner, or filter, tuned to a particular channel frequency, receives an impulse somewhat as shown at *B*, and, when this is rectified, signals shaped like *C* result. These cause relay operation *D*, which in turn operates the

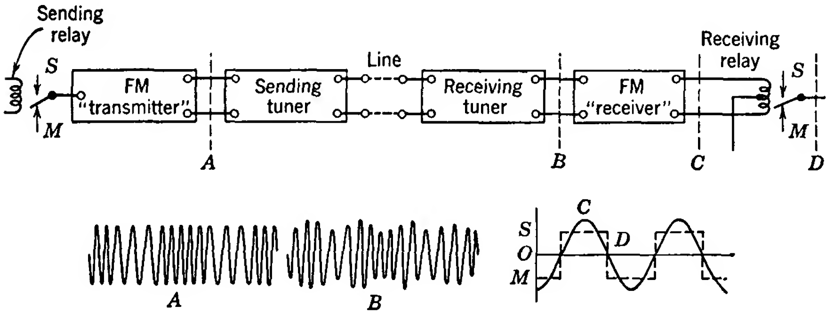


Fig. 18. Basic diagram of one channel of a frequency-modulated carrier telegraph channel. In an actual installation, many channels, such as the one shown, are connected in parallel with the telegraph transmission line. The lower figures show the approximate shapes of the currents at corresponding points in the diagram above. (Adapted from Reference 37.)

printing equipment. Other channels are connected in parallel and the signals are separated from each other by the tuners or filters. About 1,700,000 miles of carrier telegraph channels are provided in the Bell System alone. It should be mentioned that many carrier-telegraph channels can be operated over one *telephone* channel. Carrier telegraph systems have been classified as **high-frequency telegraph systems,**[34] and as **voice-frequency telegraph systems.**[35]

Frequency-Modulated Carrier Telegraph Systems.[15, 37, 38] Although amplitude-modulated carrier telegraph systems are used by both the Bell System and Western Union, the latter is standardizing on frequency-modulated systems. In these, the "carrier frequency" at which the channel operates is caused to move down 70 cycles by a marking signal, and to move up 70 cycles by a spacing signal. For this reason the terms **frequency-shift modulation** and **F–S modulation** are sometimes used.

A block diagram of a typical system is shown in Fig. 18. A sending relay arrangement such as indicated would cause the frequency-modu-

lated signal to have sudden frequency transitions and appear as in *A*. Electronic control is used, however, and the change in frequency is accomplished gradually. After removing undesired components, the transmitted and the received components are more nearly represented by *B*. This signal is passed through a **limiter,** which removes all amplitude variations, and then is demodulated in a **discriminator,** which gives out impulses such as *C*. These actuate the receiving relay. The action of the limiter and the discriminator and much information on frequency modulation will be found in Chapter 13. The Western Union has found frequency-modulation systems to be particularly free from interference from all causes.[15, 37] Typical overall frequency assignments of carrier telegraph systems are shown in Fig. 19. The manner in which one system is further broken down into individual telegraph message channels is explained in reference 15. By 1947 Western Union was operating over 500,000 miles of carrier channels. A frequency-modulation telegraph system that does not employ the conventional telegraph relay has been developed.[39]

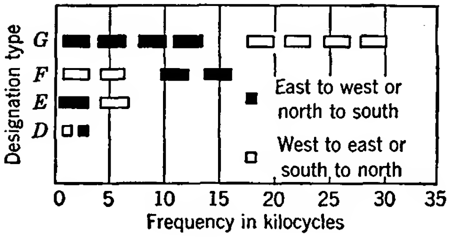


Fig. 19. Major frequency allocations of certain Western Union carrier systems. These bands are subdivided into individual message channels. (Adapted from Reference 15.)

Teletypewriter Exchange Systems.[40, 41, 42] Many industrial firms, public agencies, and similar organizations utilize teletypewriter exchange service, or **TWX,** as it is known. Subscribers to this service are provided with a teletypewriter and with a nationwide directory of other subscribers and their TWX numbers. Connections between subscribers may be made by operators or by automatic equipment.

After a connection has been established, the two parties type messages back and forth, with the decided advantage that typed copies of all transactions are available to both parties. Payment is made on a time basis. This service was established by the Bell System in 1931.

Leased Wire Service. Large industries, the press, public agencies, and other organizations, such as financial houses, are provided with leased-wire telegraph service by both the Bell System and Western Union. These circuits are usually terminated in printing telegraph equipment and may extend between offices and plants, branch offices, etc.

Leased channels may be set up for long periods or may be of such

nature that occasional switching is necessary. Special apparatus, often designed on a customer basis, is available for this purpose.[15, 16] The varioplex system, described earlier in this chapter, and **telegraph concentrators**[2, 16] are useful in providing the occasional transmission needed by TWX and leased-wire customers.

Telegraph Switching Systems. The problem of routing telegraph messages in a large office is quite involved. Most telegraph messages are very short, often consisting of only a few words, and incoming lines bring messages to points scattered throughout the country. In the past, incoming messages have been received as printed letters on tape printers. The tapes then were gummed onto paper blanks, passed to route clerks, distributed to the proper outgoing operator, and re-sent by teletype equipment.[15] What is known as **reperforator switching** in automatic switching centers is now used. With this method, a message is both typed on the tape and recorded as perforations. An operator reads the destination on the tape and by push-button control routes the message which then is automatically sent, over the proper trunk, from the perforations.*

Simultaneous Telephone and Telegraph Operation. If suitable equipment is used, it is possible to provide simultaneous telegraph and telephone service over the same line wires. This may be done by using high-frequency carrier systems as previously discussed. Other methods are the **simplex circuit** and the **composite circuit.**

Simplex Telegraphy. The method of obtaining a simplex circuit over the four wires of a phantom group is illustrated in Fig. 20. Thus, with four line or cable wires, three telephone and one telegraph channel can be operated simultaneously. If the theory of the phantom circuit presented on page 224 is studied, no discussion of the theory of the simplex circuit will be required. A simplex circuit may be obtained from, and is often used on, one pair of telephone cable wires.

Composite Telegraphy. The composite arrangement is extensively used because a telegraph channel can be obtained from *each* of the four telephone line wires of the phantom group of Fig. 21. With a composite circuit,[4] the upper telegraph frequencies transmitted are about 80 cycles per second. The inductor and capacitor combination in the telegraph circuits comprises a low-pass filter permitting the low-frequency telegraph currents to pass through to the telegraph instruments. Similarly, the inductors and capacitors in the telephone

* Methods of switching are receiving much attention; see *Western Union Technical Review*, January and July, 1948, Vol. 2, Nos. 1 and 3, and also *Elec. Eng.*, July, 1948, Vol. 67, No. 7.

circuits constitute a high-pass filter permitting the relatively high-frequency voice currents to pass through.

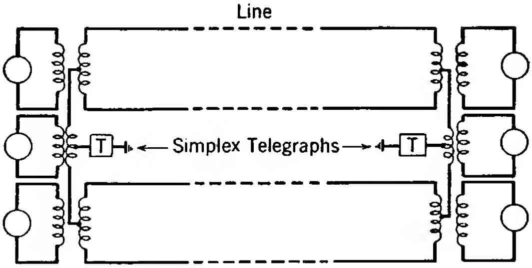


FIG. 20. A simplex telegraph circuit on a phantom group providing simultaneous telephone and telegraph service. A simplex circuit may also be provided on a pair of wires.

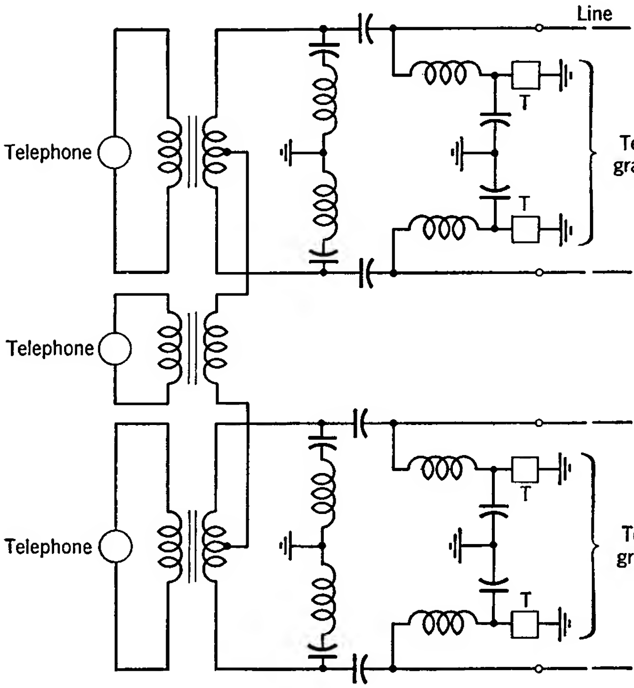


FIG. 21. Simplified circuit of terminal equipment at one end of a composited phantom group. This provides four telegraph channels, in addition to the three telephone channels, and is particularly applicable to open-wire lines.

Interference in Composited Circuits. Composite equipment is so designed that normally the telegraph signals do not interfere with tele-

phone conversations. Also, the design is such that this equipment causes very little transmission loss in the telephone circuits. The telegraph frequencies employed are not readily reproduced by a telephone receiver. The harmonics of these frequencies are in the voice range, however, and would cause interference if not suppressed by the composite arrangements. If, owing to defective or improper terminal apparatus, the telegraph currents are audible in the telephone circuit, the name **telegraph,** or **Morse, thump** is given to this interference. (See also page 318.)

The Metallic Telegraph System. There is a growing tendency among communication companies to place important circuits in cable whenever it is practicable to do so. The **metallic telegraph** system was designed especially to provide telegraph service over such cable circuits. These telegraph channels are often provided by superimposing the telegraph messages on the cable pairs already carrying telephone speech currents. The composite arrangement is common. The cable conductors are very close together, are loaded, and are provided with repeaters (page 399). Interference from the superposed telegraph would be excessive if the currents and voltages used for telegraphing were not kept far below the values used for ordinary telegraphy.[43] Line currents of 4 or 5 milliamperes are used in typical installations, these being of the same order of magnitude as the telephone currents, but of lower frequency.

Facsimile Transmission. Several facsimile methods have been developed. In some systems the message, picture, or drawing to be sent is scanned with a light beam, and corresponding electric impulses are produced with a phototube. In other systems a special paper called **Teledeltos recording paper** is used at both the sending and receiving ends. This is a coated conducting paper which is caused to become black at a point where electric current passes through it. It is not affected by light and requires no "photographic" processing.[15]

When using Teledeltos paper for transmitting information, the message is impressed on the paper with a soft lead pencil or from a special carbon paper. As the message is written, a film on the surface is ruptured, and a low resistance path through the paper is formed at each point. This makes possible an electric method of scanning and transmission.[44] A system transmitting from a sheet such as a telegraph blank is called the **Telefax.**[15] A system transmitting from a written message on a tape is called a **Teletape.**[44]

Submarine Telegraph Cable Service.[15, 45, 46] The first cables laid for submarine telegraph service were operated by hand. At the

receiving end the signals were used to deflect the mirror of a sensitive moving-coil galvanometer.[47] This receiving device was superseded by the Kelvin siphon recorder, which not only was faster but also supplied a permanent record of the message. Automatic transmitting devices were developed which operated the cables at the maximum possible speed. Magnifiers, or amplifiers, which strengthened the received currents also were used. Cable printer telegraph equipment has been developed and installed.

In illustrating the advances that have been made, one writer[48] compared the output of a cable, laid 50 years ago, with the various methods of operation applied. With the galvanometer, the output was about 70 letters per minute; with the siphon recorder, about 80 letters per minute; when duplexed, about 160 letters per minute; with automatic transmission, about 220 letters per minute; and with printing equipment, about 375 letters per minute.

Originally, most cables were operated in sections;[49] for example, between New York and London the messages were received and re-sent at Nova Scotia, Newfoundland, and Ireland. About 1918 the first successful through operation was made by connecting the sections through equipment which reshaped and amplified the signals.

Two types of amplifiers[46, 50] (magnifiers or repeaters, as they are often called) have been used. The first type consists of mechanical regenerative repeaters, including the rotary type, and the synchronized tuning-fork type. Vacuum-tube amplifiers (the second type) are superior to the mechanical devices, especially for high-speed work. The development and installation of loaded cables for submarine telegraphy (page 261) was a great advancement.

The first high-speed loaded submarine telegraph cable was laid by Western Union in 1924. The capacity of this cable was 1500 letters per minute. Another similar cable was laid in 1926, having a speed of 2500 letters per minute. In 1928 a cable was laid which had a speed of 2100 letters per minute in one direction, but which could be duplexed, giving over 2800 letters per minute. Pictures are transmitted over transoceanic cables.[51]

Submarine cables are subject to interference,[52] and this is one of the factors limiting cable-transmission speed. The interference comes from artificial sources, such as electrical power plants and railway systems in the vicinity of cable terminations, and from natural sources, for example, electrical and magnetic storms accompanying the aurora borealis. At certain times these natural disturbances may render the submarine cables inoperative.

REFERENCES

1. *Annual Report,* Western Union Telegraph Co., 1948.
2. *American Standard Definitions of Electrical Terms,* 1941, A.I.E.E.
3. PENDER, H., and K. MCILWAIN. *Electrical Engineers' Handbook,* Volume 2. Electric Communication and Electronics, John Wiley & Sons.
4. BELL, J. H. Composite Telegraphy. *Bell Lab. Record,* December, 1928, Vol. 7, No. 4.
5. HAMILTON, B. P. Single-line Telegraphy in Early Bell System Practice. *Bell Lab. Record,* November, 1945, Vol. 23, No. 11.
6. MCNICOL, A. M. *American Telegraph Practice.* McGraw-Hill Book Co.
7. *Encyclopædia Britannica,* Vol. 21, 14th edition, "Telegraphy."
8. GLAZEBROOK, R. A. *A Dictionary of Applied Physics.* Vol. 2, Electricity. The Macmillan Company.
9. BELL, J. H. Printing Telegraph Systems. *Trans. A.I.E.E.,* Vol. 39, Part I, p. 167, 1920.
10. DUSENBERRY, W. L. Teletypewriter Service and Its Present Day Uses. *Bell Telephone Quarterly,* April, 1931, Vol. 10, No. 2.
11. WATSON, E. F. General Features of Teletypewriters. *Bell Lab. Record,* October, 1938, Vol. 17, No. 2.
12. SWEZEY, B. S. A New Page Teletypewriter. *Bell Lab. Record,* April, 1939, Vol. 17, No. 8.
13. Long Lines Department, American Telephone and Telegraph Co. *Principles of Electricity Applied to Telephone and Telegraph Work.*
14. Some Features of the Telephone and Telegraph Systems, Part II. National Electric Light Association. *Publication* 144, July, 1931.
15. D'HUMY, F. E., and P. J. HOWE. American Telegraphy After 100 Years. *Trans. A.I.E.E.,* Vol. 63, 1944.
16. DUNCAN, J. A., R. D. PARKER, and R. E. PIERCE. Telegraphy in the Bell System. *Trans. A.I.E.E.,* Vol. 63, 1944.
17. NYQUIST, H. Certain Topics in Telephone Transmission Theory. *Trans. A.I.E.E.,* April, 1928, Vol. 47.
18. HERMAN, J. Signal Distortion and Morse Transmission Quality. *Bell System Tech. J.,* April, 1929, Vol. 8, p. 267.
19. WATSON, E. F. Fundamentals of Teletypewriters used in the Bell System. *Bell System Tech. J.,* October, 1938, Vol. 17, No. 4.
20. SHANCK, R. B., F. A. COWAN, and S. I. CORY. Recent Developments in the Measurement of Telegraph Transmission. *Bell System Tech. J.,* January, 1939, Vol. 18, No. 1.
21. NYQUIST, H., R. B. SHANCK, and S. I. CORY. Measurement of Telegraph Transmission. *Trans. A.I.E.E.,* February, 1927, Vol. 46.
22. SHANCK, R. B. Neutralization of Telephone Crossfire. *Bell System Tech. J.,* July, 1926, Vol. 5, No. 3.
23. KING, R. W. The Aurora Borealis. *Bell Telephone Quarterly,* April, 1926, Vol. 5, No. 2.
24. CORR, M. S. Earth Current Measurements. *Bell Lab. Record,* December, 1935, Vol. 14, No. 4.
25. CONNERY, A. F. Non-Rotary Regenerative Repeater. *Trans. A.I.E.E.,* Vol. 46, p. 897, 1927.

26. Bell, J. H. Regenerative Telegraph Repeater. *Bell Lab. Record,* August, 1930, Vol. 8, No. 12.
27. Bell, J. H. Reforming Telegraph Signals. *Bell Lab. Record,* July, 1936, Vol. 14, No. 11.
28. McCann, T. A. Developing the 106A1 Regenerative Telegraph Repeater. *Bell Lab. Record,* October, 1936, Vol. 15, No. 2.
29. Wilder, H. F. Electronic Regeneration of Teleprinter Signals. *Elec. Eng.,* January, 1946, Vol. 65, No. 1.
30. Hearn, R. B. An Electronic Regenerative Repeater for Teletypewriter Signals. *Trans. A.I.E.E.,* Vol. 66, 1947.
31. Holcomb, Jr., P. Basic Principles of the Varioplex Telegraph. *Trans. A.I.E.E.,* Vol. 60, 1941.
32. Shute, E. R. The Varioplex: A New Development in Telegraphy. *Trans. A. I. E. E.,* Vol. 60, 1941.
33. Bramhall, F. B. High-speed Voice-frequency Carrier Telegraph System. *Trans A.I.E.E.,* Vol. 59, 1940.
34. Colpitts, E. H., and O. B. Blackwell. Carrier Current Telephony and Telegraphy. *Trans. A.I.E.E.,* February, 1921, Vol. 40.
35. Hamilton, B. P., H. Nyquist, M. B. Long, and W. A. Phelps. Voice-frequency Carrier Telegraph System for Cables. *Trans A.I.E.E.,* February, 1925, Vol. 44.
36. Matte, A. L. Advances in Carrier Telegraph Transmission. *Bell System Tech. J.,* April, 1940, Vol. 19, No. 2.
37. Bramhall, F. B. Carrier Telegraph Systems. *Elec. Eng.,* August, 1944, Vol. 63, No. 8.
38. Davey, J. R., and Matte. Frequency Shift Telegraphy—Radio and Wire Applications. *Trans. A. I. E. E.,* Vol. 66, 1947.
39. Cusack, F. H., and A. E. Michon. An FM Telegraph Terminal Without Relays. *Trans. A.I.E.E.,* Vol. 66, 1947.
40. Singer, F. J. The Teletypewriter Exchange Network. *Bell Lab. Record,* January, 1938, Vol. 16, No. 5.
41. Knowlton, A. D., G. A. Locke, and F. J. Singer. Switchboards and Signaling Facilities of the Teletypewriter Exchange System. *Bell System Tech. J.,* October, 1936, Vol. 15, No. 4.
42. Pierce, R. E., and E. W. Bemis. A Transmission System for Teletypewriter Exchange Service. *Bell System Tech. J.,* October, 1936, Vol. 15, No. 4.
43. Bell, J. H., R. B. Shanck, and D. E. Branson. Metallic Polar-Duplex Telegraph Systems for Cables. *Trans. A.I.E.E.,* Vol. 44, p. 316, 1925.
44. Pollard, L. G. Western Union Teletype Facsimile. *A.I.E.E. Technical Paper* No. 48–74, December, 1947.
45. Pratt, H., and J. K. Roosevelt. Developments in the Field of Cable and Radio Telegraph Communications. *Trans. A.I.E.E.,* Vol. 63, 1944.
46. Cramer, C. H. Some Modern Techniques in Ocean Cable Telegraphy. *Elec. Eng.,* August, 1947, Vol. 66, No. 8.
47. Clokey, A. A. Automatic Printing Equipment for Long Loaded Submarine Telegraph Cables. *Bell System Tech. J.,* July, 1927, Vol. 6.
48. Angel, H. Printing Telegraphs on Non-Loaded Ocean Cables. *Trans. A.I.E.E.,* Vol. 46, p. 884, 1927.
49. Coggeshall, I. S. Submarine Telegraphy in the Post War Decade. *Trans. A.I.E.E.,* April, 1930, Vol. 49, p. 476.

50. CURTIS, A. M. The Application of Vacuum Tube Amplifiers to Submarine Telegraph Cables. *Bell System Tech. J.*, July, 1927, Vol. 6.
51. MILNOR, J. W. Picture Transmission by Submarine Cable. *Trans. A.I.E.E.*, Vol. 60, 1941.
52. GILBERT, J. J. Extraneous Interference on Submarine Telegraph Cables. *Bell System Tech. J.*, July, 1926, Vol. 5.

REVIEW QUESTIONS

1. What is meant by the terms single-current and neutral telegraph systems?
2. Explain the difference between open-circuit and closed-circuit systems.
3. What are the average and maximum rates of manual, or Morse, operation?
4. If perfectly rectangular impulses such as shown in Fig. 3 are transmitted, what will be the maximum alternating frequency components present?
5. About what are the maximum frequency components transmitted in an average telegraph channel?
6. In modern practice what type of relays are used in single-current systems?
7. Explain the operation of the double-current or polar telegraph system.
8. Give several reasons why the polar system is preferable.
9. What two types of duplex systems have been developed, and what type usually is employed at present?
10. Explain the basic principle of the start-stop teletypewriter.
11. Explain the basic principle of the multiplex system.
12. Could a hand-operated teletypewriter operate directly into a multiplex system?
13. Why are perforated tapes used with multiplex operation?
14. What is meant by the term telegraph channel? Name several systems that provide multichannel operation over one grounded line.
15. Briefly discuss the two possible bases for analyzing telephone transmission. Which is usually employed?
16. What is meant by the word baud?
17. If 60 words per minute will produce a third harmonic of about 75 cycles, why are wider channels sometimes provided for telegraph service?
18. Why is bias distortion so very important in telegraphy? Does this have any bearing on the use of the polar system?
19. Name and distinguish between the several types of distortion.
20. What two functions are performed by regenerative repeaters? What types are used?
21. Briefly explain the basic principle of the varioplex system. Will it operate into multiplex equipment? Into carrier telegraph systems?
22. What types of carrier telegraph systems are used? Which type is most widely used in the Bell System? In Western Union systems? Can you give reasons for this?
23. Describe the amplitude-modulation telegraph system.
24. Describe the frequency-modulation telegraph system.
25. Describe the teletypewriter exchange system.
26. What is meant by reperforator switching?
27. Briefly describe the means by which simultaneous telephone and telegraph service are given over the same circuits.
28. Why is the so-called metallic telegraph system of increasing importance?

29. Explain the difference between the telefax and teletape systems.
30. Briefly describe methods of giving transoceanic telegraph service over cables.

PROBLEMS

1. Referring to the circuit of Fig. 22, calculate the magnitudes and indicate the directions of the current flowing in each part of the circuit with the keys at

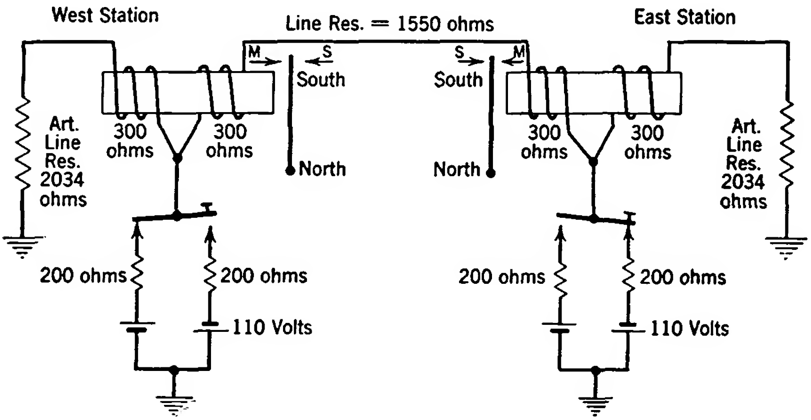


FIG. 22. Simplified differential duplex system showing typical resistance values. (Adapted from Reference 14.)

each station in the positions shown. Determine whether each relay armature, with the magnetic polarities indicated, will rest on the marking (m) or the spacing (s) contacts.
2. Repeat Problem 1 for each of the four possible positions of the sending and receiving keys, drawing the equivalent circuits and indicating the magnitudes and directions of the currents on these circuits, and also showing the resulting relay positions.

CHAPTER 10

TELEPHONE EXCHANGE SERVICE AND SYSTEMS

Introduction. Telephone service is of two types, **exchange service** (often called **local service**), and **toll service** (also called **long distance service**). This chapter will consider exchange service and systems.

A **telephone exchange**[1] is a "telephone system for providing telephone communication within a particular local area, usually within or embracing a city, town or village, and environs." Service within an exchange is usually rendered at a fixed monthly rate or on a metered basis.

In a telephone system, the various lines are switched at the **central office,** defined[1] as "an office in a telephone system providing service to the general public where orders for or signals controlling telephone connections are received and connections established." A telephone exchange serving a small community may have only one central office and thus is a **single-office exchange.**[1] For large towns and cities, more than one central office is needed, requiring a **multioffice exchange.**[1] The offices just mentioned are **local central offices,** defined[1] as "a central office arranged for terminating subscriber lines and provided with trunks for establishing connections to and from other central offices."

The **trunk**[1] mentioned in the preceding paragraph is a "telephone line or channel between two central offices or switching devices, which is used in providing telephone connections between subscribers generally." By means of **interoffice trunks,**[1] a call that originates in one central office can be passed to another office for completion.

The telephone customer or **subscriber**[1] is provided with a **telephone set.**[1] A **telephone station**[1] is a telephone set and associated wiring and apparatus installed on the subscriber's premises. A telephone set is sometimes called a **subscriber set** or **subset.** Telephone sets are of two general types, local-battery telephone sets, and common-battery telephone sets.

Simple Telephone Systems. A simple two-way telephone circuit is shown in Fig. 1. This circuit is objectionable because of the losses offered by the unused transmitter and receiver to the direct-current

component for the transmitters and to the alternating-current speech component. Transmitters and receivers are often connected in series for **interphones** such as those used in apartment houses.

Two transmitters and two receivers could be connected in parallel to a line and battery as indicated in Fig. 2. This arrangement will give two-way conversation, but it also is inefficient. Under some conditions[2] its use may be more efficient than that of Fig. 1.

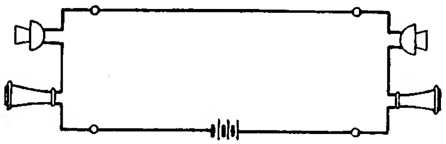


FIG. 1. A simple series telephone circuit.

Variable and Invariable Substation Circuits. Satisfactory telephone communication could be carried on over a system in which the receiver was removed from the circuit at the sending end when speaking was in progress and in which the transmitter was removed from the receiving end when listening was going on. Such an arrangement would be efficient because it would eliminate the losses in the other units not being used. Such a circuit is known as a **variable telephone circuit.** A circuit in which such changes cannot be made is an **invariable telephone circuit.**[2]

To eliminate troublesome switching operations, the invariable telephone circuit is commonly employed. There is consequently a loss of power in the receiver of a set during talking, and also a loss of power in the transmitter of a set during listening.

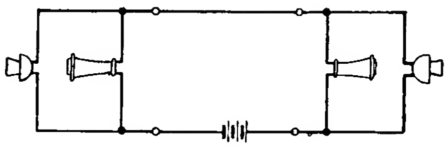


FIG. 2. A simple parallel telephone circuit.

Types of Telephone Systems. Classified on the basis of the method of switching the telephone subscriber lines, telephone systems are either **manual systems** or **dial systems.** In the manual system the switching is done by **telephone operators.**[1] These systems will be considered in the first part of this chapter, and dial systems will be considered in the last part.

Manual telephone systems are of two general types, **local-battery systems** and **common-battery systems.** A hand-operated generator or **magneto** is used in signaling in the local-battery system; hence, the designation **magneto telephone system,** or **magneto local-battery telephone system.**

Magneto Local-Battery Telephone Sets. In early telephone systems a **local battery** was installed at each telephone station to supply

the direct current for the telephone transmitter, and a magneto was provided for signaling. These sets were very satisfactory and are often used today, particularly for rural lines, forest-service systems, and similar purposes. A typical circuit is shown in Fig. 3.

The magneto is provided with a switch which removes it from the circuit and thus prevents it from introducing a loss when it is not in use. The switch is engaged when the handle is turned. The ringer is of high impedance and is usually connected or "bridged" across the line at all times.

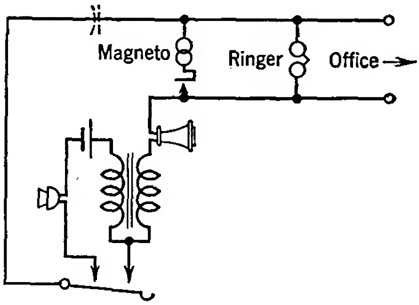


FIG. 3. Local-battery, or magneto, telephone set.

When the receiver is removed from the switch hook, the contacts at the switch hook are closed and direct current flows through the transmitter. When sound waves strike the transmitter diaphragm, a pulsating current as shown in Fig. 4 flows through the primary of the induction coil and induces an alternating voltage in the secondary circuit. A corresponding alternating current will thus flow through the line, reproducing the sound waves at the distant station.

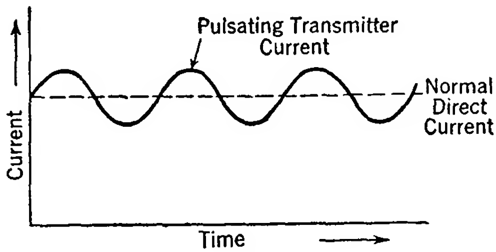


FIG. 4. Pulsating current flowing through a transmitter and associated circuit when a transmitter is excited with a pure tone.

The **induction coil** (transformer), which apparently was introduced at about the same time[3] by both Berliner and Edison, greatly improves telephone transmission. In particular: (*a*) It separates the transmitter and receiver currents so that the direct current for the transmitter does not pass through the receiver. (*b*) The induction coil may have other than a one-to-one ratio, thus "stepping up" the voltage impressed on the line, or, in other words, the coil matches the impedance of the transmitter to the higher impedance of the line. (*c*) The use of even a one-to-one ratio repeating coil will greatly increase the *percentage* change in resistance and thus increase the useful alternating-current component. For instance, if

the transmitter is in series with a battery, a receiver, and the line, the percentage change in the resistance of the circuit due to the variations in transmitter resistance will be smaller than if the transmitter is connected in series with only the primary of the induction coil and the battery.

A capacitor is often connected as shown by the dotted lines in Fig. 3, especially where a number of local-battery sets are on the same line. This is known as a "sure-ring condenser," and it prevents a set with a receiver off the hook from shunting the ringing currents. Thus, on lines where a number of sets are connected, if one party leaves the receiver off the hook the low impedance of the set will prevent another person from ringing a third party on the line. With a sure-ring condenser, however, the low-frequency ringing currents are not shunted because of the high reactance.

Line Connections of Magneto Telephone Sets. The telephone sets are connected in parallel or **bridged** across the line. Ten sets on a given line is regarded as the maximum number for good service, although as high as twenty or more are sometimes used.

From Fig. 3 it can be seen that with the telephone sets bridged across the line there will be a corresponding number of bells in parallel across the line at all times. It is important that these bells offer a high impedance to the talking currents and that they have the *same* impedance. The ringing currents, of 15 to 20 cycles (depending on how rapidly the magneto is turned), readily pass through the coils and actuate the bells.

The grounded telephone line is sometimes used in rural areas and forest telephone systems, but not so extensively as in the past. Only one wire is used, the earth acting as the other side, or the "return" for the voice currents. Grounded circuits are very susceptible to inductive interference (Chapter 14).

Some rural and forest lines use iron wire for the conductors. Iron wire has, however, a much greater resistance than copper. Also, owing to the high permeability compared to copper, the skin-effect losses (page 215) and the resulting attenuation are greater in iron than in copper wires. A *very* small percentage of telephone line conductors are iron, copper usually being used.

Polarized Ringers or Bells. The polarized ringer or bell has been used very extensively from almost the beginning of telephony. In Fig. 5(*a*) and (*b*), simplified schematic views are shown.

In these figures, c and c_1 are soft-iron cores connected at the upper end by a soft-iron strap s–s_1. This strap also serves to hold the gongs g and g_1. The soft-iron armature a–a_1 is pivoted as indicated and has attached to it the ringing device. Two high-impedance windings w

and w_1 are placed on the cores as indicated. The resistance of these windings is of the order of from one thousand to several thousand ohms, and the impedance is very much higher, depending on the frequency of the current. A permanent magnet m–m_1 is placed as shown.

There are, therefore, two magnetic circuits. The first of these is for the flux produced by the ringing current through the coils and is along the soft-iron core c, through the soft-iron strap s–s_1, through the other core c_1, across the air gap to the armature a_1, along the armature to a, and then back across the air gap at a to the core c. The other path is for the magnetic flux produced by the permanent magnet m–m_1. This path is from the lower end of this permanent magnet across the large air gap to the center of the armature a–a_1 where the flux divides, half of it passing through the armature and across the air gap at a, and the other half passing through the armature and across the air gap at a_1. Each half of the flux produced by the permanent magnet then passes through the coil c or c_1, and through the strap s–s_1 to the end of the permanent magnet.

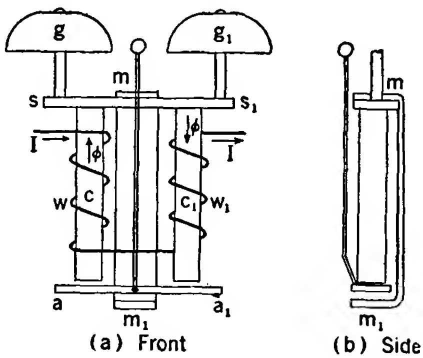


FIG. 5. A polarized ringer.

The operation of the bell can be explained in the following manner. Suppose the magnet is poled so that it gives the ends of the armature a–a_1 a south polarity. With no current flowing, no unbalance will exist and the armature will occupy the balanced position shown. When an alternating ringing current flows through the bell, flux will be produced by the windings. Assume that the current at a particular instant has the direction indicated by the arrow; then, the flux ϕ produced by this current will have the directions indicated and will tend to make the end of core c_1 a *stronger* north pole, and c a *weaker* north pole. The armature will therefore be unbalanced and will be attracted to the pole c_1, and the bell will ring. When the opposite part of the alternating current passes through the winding, the action will be reversed and the bell will again ring.

Magneto Local-Battery Switchboard. When one telephone user wishes to talk with another on the *same* line of a magneto local-battery system, the first will call the second party by ringing with the magneto. This signal will be sent out before either receiver is removed from the hook, and thus all the parallel or bridged bells on the line are operated.

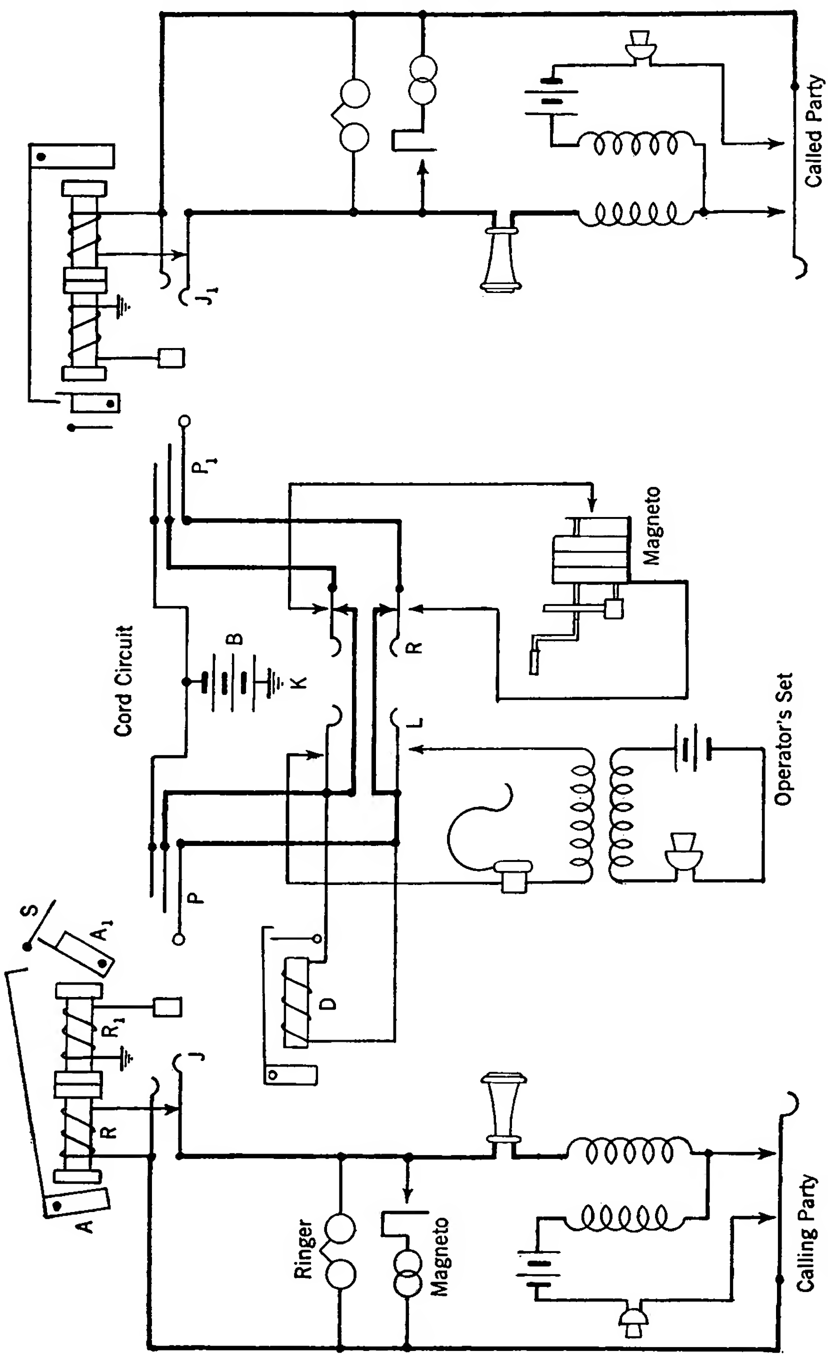


FIG. 6. Connection of two local-battery sidetone telephone sets through a typical local-battery or magneto switchboard.

Coded signals are used to call the desired party. When a magneto telephone user wishes to talk with another on a *different* line, the connections are made through a magneto switchboard (such as Fig. 6) in the central office.

To call the operator, the calling party turns the magneto crank several times and then removes the receiver from the hook. The magneto sends out an alternating current on the line and through the relay R. This energizes the relay, causing it to attract the armature A, which trips the armature A_1, allowing it to drop forward and raise the shutter S, which displays to the operator the number of the calling line shown on A_1. The operator answers this incoming call by inserting the plug P of one of the available vacant switchboard cords into the jack J of the calling line.

When the plug is inserted, the connection of the relay R to the other side of the line is broken, and the talking current is not shunted through the relay windings. This plug also completes the battery circuit shown, and the relay winding R_1 is energized by the battery B. The armature A_1 is then attracted to its original position and the signal is cleared. In some boards this drop must be restored manually by the operator.

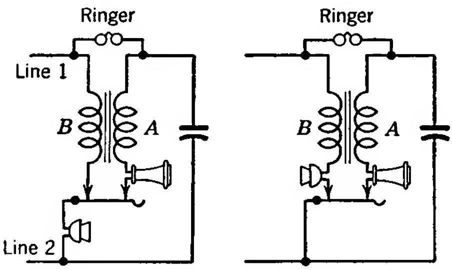
After inserting the plug P of the **cord circuit** into the jack of the calling line, the operator throws the key K of the cord used to the (left) listening position L. This connects the operator's telephone to the cord circuit and to the calling party, and the operator takes the number of the party desired. The operator then inserts the plug P_1 on the other end of the cord circuit into the jack J_1 of the called line and throws the key K to the (right) ringing position R. This disconnects both the operator's set and the line of the calling party from the circuit, and the called party is connected and can be rung by turning the central-office magneto. In some instances two separate keys are provided for listening and ringing, and in the larger magneto systems a ringing machine supplants the hand generator or magneto. When the conversation is completed, either party signals the operator by ringing on the line. This causes the disconnect drop D to operate, indicating to the operator that the connection may be cleared.

Common-Battery Telephone Sets. In the common-battery telephone system, one battery at the central office supplies the direct current to *all* transmitters in use at a given time; hence, the term **common battery.** The history[3] of this type of set is of interest, many circuits, some of questionable merit, having been developed.[2, 3, 4, 5]

Common-battery telephone circuits may be classified as **sidetone circuits** or as **antisidetone circuits.** With the first type, considerable

energy is transferred from the transmitter to the receiver, and thus a sound actuating the transmitter is plainly heard in the associated receiver as sidetone. The antisidetone circuit greatly reduces this effect.

Sidetone Circuits. A typical common-battery sidetone circuit is shown in Fig. 7(*a*). The transmitter receives direct current from the battery in the central office over the two line wires. When sound waves strike the transmitter diaphragm, the current is caused to vary in accordance with these sound waves (Fig. 4). The alternating voltage generated at the transmitter will force a current through the series path consisting of the receiver, winding A of the induction coil, and the capacitor. In a typical telephone set, this capacitor neutralizes, to some extent (depending on the frequency), the inductive reactance of the receiver, thus lowering the impedance and increasing the current flow in the receiver circuit. This current produces sidetone as previously mentioned. In addition, this current induces an electromotive force in the winding B of the induction coil (which has an unequal ratio of turns and thus acts as a step-up transformer), and this induced voltage adds to the alternating voltage produced at the transmitter terminals, hence *increasing* the total voltage impressed on the line.



(*a*) Sidetone circuit (*b*) Sidetone-reduction circuit

FIG. 7. Connections of common-battery sidetone telephone sets.

With the sidetone-reduction circuit of Fig. 7(*b*), the direct current received by the transmitter is about the same as in the sidetone circuit, and thus the alternating voltage output corresponding to the impinging sound waves is about the same. This voltage causes a current to flow through winding B of the induction coil, and this current induces a voltage in winding A. Since the induction coil now acts as a step-down transformer, the induced voltage in A is considerably less than the voltage impressed across the receiver circuit in the sidetone set. A smaller current flows through the receiver under these conditions, *and less sidetone is produced.* The *receiving* efficiency is increased but the sidetone-reduction circuit does not impress so high a voltage on the line as the sidetone circuit and is less efficient *in transmitting.*

Antisidetone Circuits. The sidetone circuit just discussed has desirable characteristics such as high transmitting efficiency. Nevertheless, noises actuating the transmitter, and amplified by it (page 94), are heard in the receiver and interfere with conversation. Also, if the telephone user hears his voice loudly reproduced by the receiver when he speaks, he involuntarily lowers his voice,[6] thus reducing the useful output.

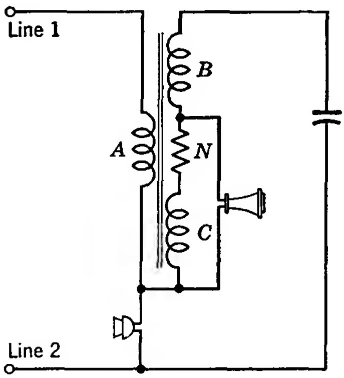


FIG. 8. An antisidetone telephone circuit with network C–N for reducing sidetone.

Several antisidetone sets have been developed. One problem is to design a circuit that is effective in largely preventing sidetone, yet is efficient in transmitting and receiving.

Antisidetone Circuit Using a Balancing Network.[6, 7] The balancing network of Fig. 8 is used in several antisidetone sets. It is the circuit of Fig. 7 with the addition of the balancing network C–N.

In the circuit of Fig. 7(a) the transmitter forces speech currents through coil B, and also through coil A and the receiver, this current causing sidetone. Coil C of Fig. 8 is wound on the magnetic core of the induction coil (transformer) in such a direction that the voltage induced in it tends to cause speech currents to flow through the receiver in the direction opposite to the flow of Fig. 7(a). This cancellation reduces sidetone. The resistor N, which may be the resistance of coil C, is effective in producing the correct phase relations in the circuit so that the sidetone cancellation is effective over a wide frequency range.

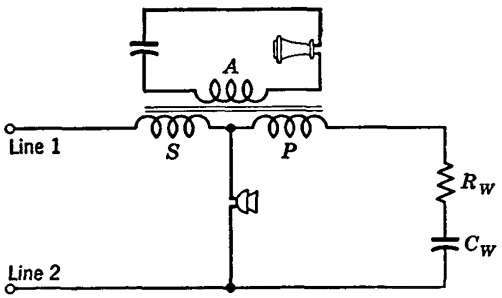


FIG. 9. Simplified connections of an antisidetone circuit using a hybrid coil, or bridge transformer.

Antisidetone Circuit Using a Hybrid Coil. The theory of the hybrid coil is considered on page 401. Briefly, the circuit of Fig. 9 is so adjusted that the output of the transmitter divides, and, for the correct values of C_w and R_w (which may be the resistance of coil P),

The Western Electric handset. (Courtesy Bell Telephone System.)

the speech currents flowing through coils P and S are equal. Since these currents will also be opposite in direction, the output of the transmitter produces (theoretically, and at a given frequency) no induced voltage in coil A, and no sidetone.

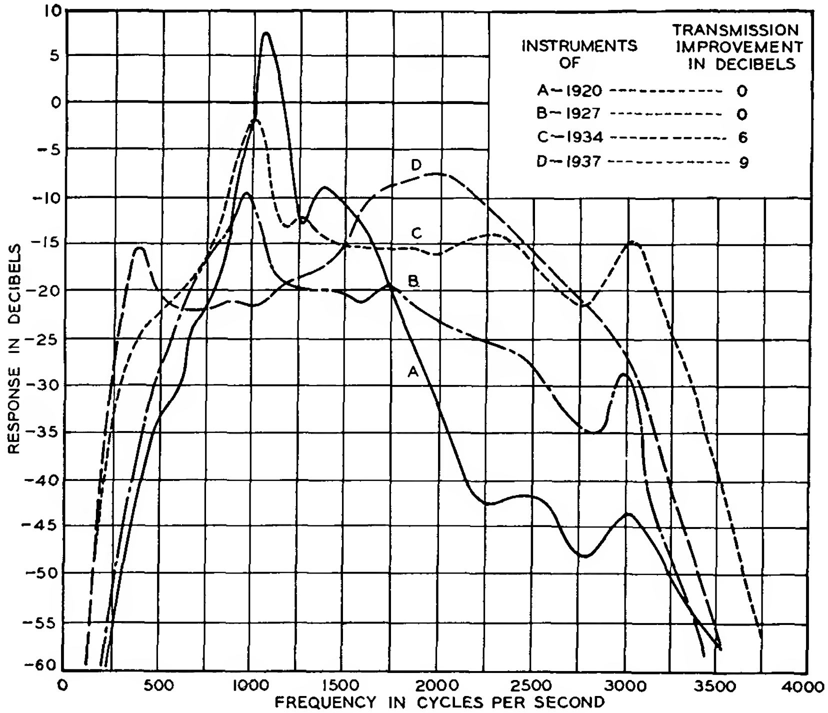


FIG. 10. Comparison of the response-frequency characteristics of telephone sets. Curve D is typical of the sets now (1949) in use. (Reference 11.)

Modern Telephone Sets. The transmitters and receivers used in modern telephone sets were discussed in Chapter 4. As was explained, much progress has been made in the design of transmitters and receivers.[7, 8, 9, 10]

Antisidetone circuits are used in modern telephone sets. The capacitors, inductors, and other components are of superior design. The performance of modern telephone sets is similar to that of curve D of Fig. 10. Most new telephones are of the **handset type.**[1]

In early handsets, the mechanical coupling afforded by the handle to the transmitter and receiver was an important factor in causing sustained oscillations or "howling." In certain early European handsets this difficulty was met by using insensitive transmitters and

receivers so that howling did not develop. In the design of the modern handset, howling because of mechanical coupling is controlled.[9] But, even with the antisidetone circuit, *some* of the output from the receiver does reach the transmitter through the air, and this acoustic coupling is a very important factor limiting the sensitiveness of the set. New telephone sets, somewhat revolutionary in nature, are now (1949) in the practical trial stage.

Line Connections in Common-Battery Systems. Individual, two-party, and four-party services are provided by most common-battery systems. That is, a telephone user can obtain service over a line on which no other sets are connected (**individual line**[1]), or over a line used jointly with one other customer (a **two-party line**), or over a line used jointly with three others (a **four-party line**). For talking, the telephone sets are connected directly between the two wires of the **subscriber line**[1] (sometimes called the **subscriber loop**) leading to the central office.

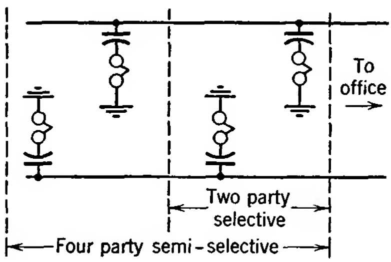


FIG. 11. Methods of ringing to ground giving the services indicated. The talking equipment is connected between wires.

The ringers (or bells) are *not* connected directly between line wires, the arrangement depending on the particular system and the type of ringing provided. In the **selective ringing** method used on two-party and four-party lines, the ringer of the called subscriber's station *only* is rung.[1] In the **semiselective ringing** method, the ringers of two subscribers' stations are rung, differentiation between subscribers being by a one-ring, two-ring code.[1]

Two-Party Selective, Four-Party Semiselective Systems. One of the most common methods of connecting the subsets to the line for ringing is illustrated by Fig. 11. This system provides two-party selective ringing and four-party semiselective ringing. Two bells are connected between one line wire and ground, and two are connected between the other line wire and ground. It is possible to ring two bells over this ground connection without ringing the other two. In a four-party connection of this type, a coded signal is used, and each party hears the other party's signal.

Four-Party Selectivity, Biased Ringers. If the armature of a polarized ringer (page 348) is held against one of the electromagnet cores by a small spring, the ringer is said to be a **biased ringer,** and

it will respond only to pulsating current of one polarity. A pulsating current in one direction will ring the bell; one in the opposite direction will merely attract the armature more strongly to the pole piece. If now the ringer is biased with a spring on the opposite side so that the armature is held against the opposite pole piece, then it will ring on a pulsating current of the opposite polarity.

It might be inferred that, if biased ringers were used in Fig. 11, four-party selectivity would be obtained, but such is not the case. To prevent the flow of direct current, which would operate central-office supervisory equipment, the ringers must be connected in series with capacitors, and these convert the pulsating currents into alternating currents and eliminate the polarity feature.

The connections of Fig. 12 are used with a biased ringer to obtain four-party selectivity. A sensitive high-impedance relay is connected in series with a capacitor and across the line at each of the four stations. Two oppositely biased ringers are connected between each line wire and ground through contacts which normally are open, but which may be closed by the relay. Pulsating ringing voltage is impressed between one side of the line and ground, with the other line wire grounded. The ringing current causes the armatures of each relay to close the contact, and then the properly biased bell rings. The relays open as soon as the ringing current ceases. It is evident that this system requires considerably more apparatus and maintenance than the semiselective system previously described.

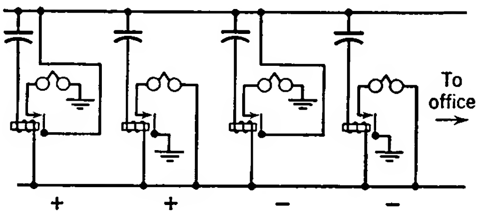


FIG. 12. A method of giving four-party selective ringing. Plus and minus signs indicate the polarity of the currents to which the biased ringers respond.

From the standpoint of inductive interference between paralleling power lines and telephone circuits, it is desirable to have as high an impedance as possible between the telephone line wires and ground, and also to have these impedances the same. Grounded ringers in grounded two-party and four-party semiselective ringing may, under bad power exposure conditions, cause noisy lines (page 549). In some instances, where there is a high direct ground potential, it may be necessary to have the four-party selective ringers connected between the wires instead of to ground, and the relays and capacitors connected to ground instead of across the line.

Biased ringers used on systems having alternating ringing current prevent tapping of the bells due to inductive or other electrical disturbances.

Cold-cathode gas tubes (page 277) are used in some instances instead of relays to make possible selective ringing on four-party lines.[12, 13] As shown in Fig. 13, the tubes are connected with a resistor of about 100,000 ohms in the starter, or control anode, and the main anode is connected in series with the ringer. When party *A* is to be rung, line 1 is grounded and line 2 is made negative. This causes the control gaps at *A* and *C* to break down. The main gap at tube *A* also breaks down (page 277), and the alternating component of the pulsating ringing current causes ringer *A* to operate. The main gap

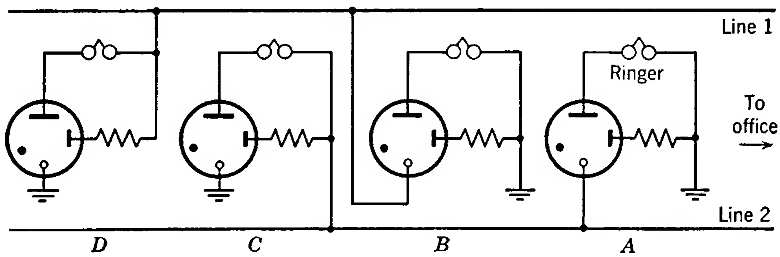


FIG. 13. Connections of ringers and cold-cathode gas triodes for giving four-party selective ringing. The small circular electrode is the cold cathode. The small straight electrode is the starter, or control, anode, and the large straight electrode connected to the ringer is the main anode. The black dot indicates that the tube contains gas.

at *C* does not break down and pass current because the main anode is negative with respect to the cathode. To cause ringer *C* to respond instead of *A*, line 2 is made positive, rather than negative. For ringing *B* or *D*, line 2 is grounded and negative pulsating ringing voltage is placed on line 1 to operate ringer *B*; positive pulsating ringing voltage is placed on line 2 to operate ringer *D*. This system offers many advantages over the relay method of providing four-party selective service.[12]

Four-Party Selectivity, Harmonic Ringing. The harmonic system of providing selective service is widely used in the United States on common-battery circuits. The harmonic ringer has an armature tuned mechanically to the desired natural period of vibration. The armature and clapper rod are supported by stiff springs; and, by varying the thickness of the spring used and the mass and the position of the weights on the clapper rod, the ringers can be made to respond to

the frequency desired. In one system, one bell rings on a current of $16\frac{2}{3}$ cycles per second; another, on $33\frac{1}{3}$ cycles; another, on 50 cycles; and the fourth, on $66\frac{2}{3}$ cycles. The bells are connected through capacitors and directly across the line. These frequencies are generated with rotating equipment.

It is evident that all these currents are multiples of the $16\frac{2}{3}$-cycle current. There is accordingly a strong tendency for each ringer to respond to the current for the others and to ring or buzz lightly. The development of the vibrating ringing-current generator made practicable a system in which the frequencies were not multiples. One of these employs frequencies of 30, 42, 54, and 66 cycles per second. The fact that four-party selective service can be provided by a harmonic system *without a ground connection* is one of its greatest advantages. Electronic devices have been used to a limited extent to generate the harmonic ringing impulses.

Subscriber Lines. These lines (or loops) connect the subscriber station equipment and telephone set to the central office. A telephone drop[1] of insulated paired wires extends from the terminals of the interior wiring of the telephone station to the aerial cable box. The various aerial cables are brought together and spliced into larger cables leading to the central office. These cables terminate in an "underground" cable box where the aerial and underground cables are connected, usually through suitable protection to prevent the exposed aerial cables from bringing hazardous currents or voltages into the underground cables (page 568). If the cable terminals are a considerable distance from the location of the telephone set, a few spans of iron open wire are often installed instead of making the entire connection of paired conductors.

The cables (page 250) used are paper insulated and have lead or plastic sheaths (page 236). The development of exchange cable is shown in Table I. Experimental work[14] has been done on having 3030 pairs of No. 28 A.W.G. wires in a standard $2\frac{5}{8}$-inch outside diameter sheath. Cables having such small conductors can be used only in congested areas where the subscriber lines are short. For longer lines, larger conductors must be used, so that the transmitters will receive sufficient direct current from the central office and so that the losses to the speech currents will not be too large.

Common-Battery System—Non-Multiple Switchboard. The common-battery, non-multiple switchboard is suited for either a village or a **private branch exchange (P.B.X.)** in a large store, hotel, or similar establishment. In this system, the line jacks of *each* party served by the switchboard are within reach of *each* operator. The

TABLE I

PAIRS OF WIRES IN EXCHANGE CABLES

Year	No. 19 Gauge	No. 22 Gauge	No. 24 Gauge	No. 26 Gauge
1892	100			
1895	152			
1896	208			
1901	303	404		
1902	...	606		
1912	...	909		
1914	...	...	1212	
1918	455			
1928	...	...	....	1818
1939	...	...	....	2121

number of lines that can be handled is limited by at least two factors: *first,* an operator can handle only a certain number of incoming calls per unit of time; and *second,* an operator can reach only a given switchboard area without moving from her chair.

To explain the operation of Fig. 14, suppose that party *A* wishes to call *B*. To attract the attention of the operator, *A* will remove the receiver from the hook, which will close the switch-hook contacts and allow direct current to flow through the line relay *LR*. When this relay is energized it will pull up the armature closing the path through the line lamp *LL* and thus signal the operator.

The operator will answer this signal by inserting the answering plug of a cord circuit into the line jack *LJ* of the calling party which is indicated by the lighted line lamp. The insertion of this plug does three things. *First,* the contacts will be spread apart, de-energizing the line relay *LR,* allowing the armature to fall back due to gravity, and thus extinguishing the line lamp. *Second,* the tip contact T and the ring contact *R* of the plug will make contact with *T* and *R* of the line jack. The closing of these two contacts will permit direct current from the common battery *CB* to flow out to the calling telephone *A*. The path of this current will be from the positive battery (grounded) terminal through one half of the inductive supervisory relay *SR* to the tip of the plug, then, through the telephone set, back through the ring of the plug, through the other half of the supervisory relay *SR,* and to the negative battery terminal. The current through the two halves of the supervisory relay *SR* will attract the armature, opening the contact *C*. *Third,* the sleeve contact *S* on the plug will touch the ground contact *S* on the jack. The supervisory lamp *SL* would now light if the supervisory relay had not previously opened. The circuit of the supervisory lamp is, however, now in a condition to

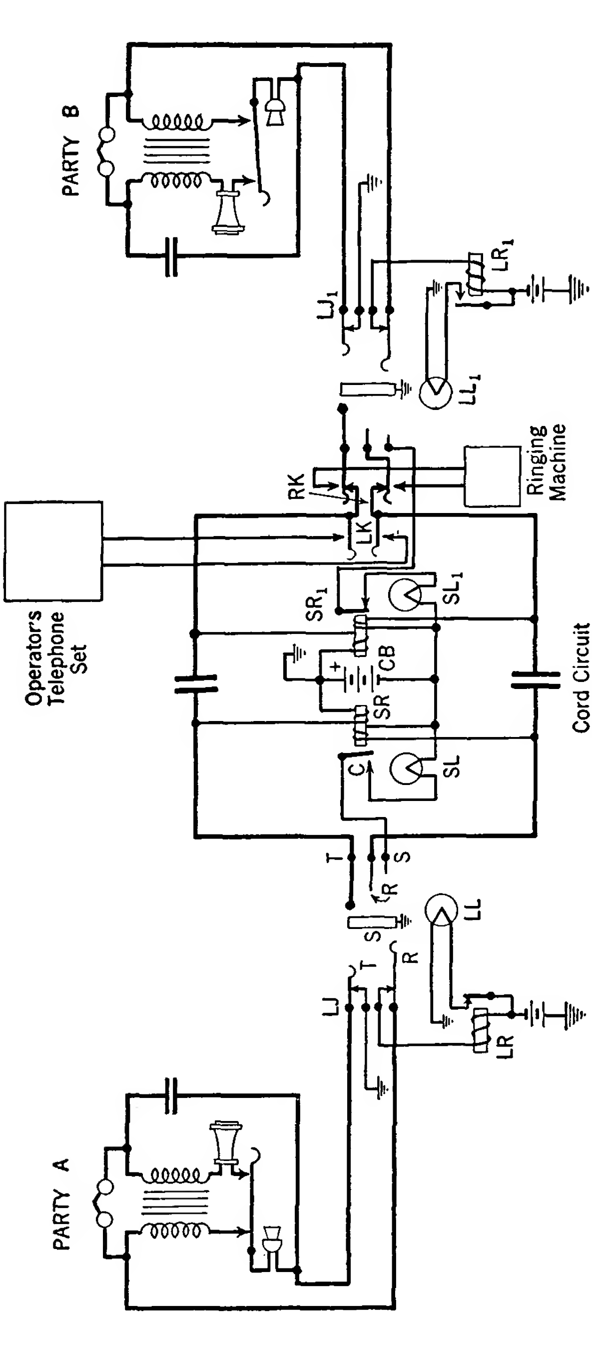


FIG. 14. Simplified circuit of a non-multiple, common-battery switchboard. Although sidetone telephone sets are shown, antisidetone sets may be used.

light the lamp as soon as this relay contact closes, as it does in a later operation.

The operator then throws the listening key *LK* to the listening position. This closes the contacts at the key and connects the operator's telephone set across the line. The desired number is obtained from the calling party, and, if the called line is not busy, the operator inserts the plug of the calling cord corresponding to the answering cord previously used into the line jack of the called party.

When this plug is inserted, the contacts of the line jack LJ_1 are opened, and therefore when the receiver of the called party B is removed from the hook the line lamp LL_1 does not light. The insertion of the plug does, however, light the supervisory lamp SL_1 through the grounded battery and ground on the sleeve. It should be remembered that the receiver at B is still on the hook and that no current is flowing through SR_1 to hold open the armature contact of this relay. The operator now throws the ringing key *RK* to the ringing position. This action connects the called line to the ringing supply and opens the contacts to the line of the calling party so that the calling party will not receive the unpleasant ringing current. The operator will continue to ring until the called party responds.

When the called party B removes the receiver from the hook, the contacts at the receiver are closed and current from the battery flows out to the called set B through the two halves of the supervisory relay SR_1. This current pulls up the armature of this relay, thus opening the contact and extinguishing the supervisory lamp SL_1, indicating to the operator that called the party has answered and that ringing may be discontinued. The two parties A and B then converse. Attention is called to the fact that all the switchboard lamps associated with this switchboard circuit are dark when conversation is in progress.

When A and B have completed their conversation they replace the receivers on the hooks. This opens the receiver-hook contacts at both of the sets and breaks the path of the currents through SR and SR_1, respectively. The armatures of these relays drop back, closing the associated contacts and permitting current to flow through each of the lamps SL and SL_1. The lighting of these lamps indicates to the operator that the conversation is completed, and the connection is accordingly cleared by removing each plug from the jack. After the conversation the operator can be called by *slowly* moving the receiver hook up and down, causing the supervisory lamps SR or SR_1 to flash if the connection has not been cleared, or causing the line lamp to flash if the connection has been removed.

Common-Battery System—Multiple Switchboard. In the previous section the *non-multiple,* common-battery system was considered.

As explained, in this system the various lines are connected to the board only through an individual line jack. It was mentioned that this type was limited to small cities or to private branch exchange installations. One operator can effectively handle only about 150 to 200 lines, depending on the make of the board and the type of service given. Thus, when the number of incoming lines approaches this amount, a second position must be provided, or if required, a third position added. By reaching in front of each other, the operators can complete each incoming call. If more than three positions are required, however, this method would be unsatisfactory, and transfer circuits (or interposition trunks) would be necessary between the different positions.

The multiple switchboard was early designed[3] to eliminate this trunking or transfer problem. With this type of board, each operator receives certain calls and *completes all calls for parties connected to that central office.* Each operator must therefore be able to reach calling jacks connected to each line in that central office area. Accordingly, in a central office serving 10,000 users, *each operator must be able to reach 10,000 calling jacks.* An operator can, by reaching in front of the two adjacent operators, cover three positions, and thus each line must terminate in *multiple* (or parallel) calling jacks at every third position down the board. As an illustration, if the party connected to line 250 wishes to talk with the party on line 5050, the first party will be answered by the operator at whose position the answering jack for line 250 appears. This operator then completes the call by selecting from the multiple line 5050 and connecting line 250 to it.

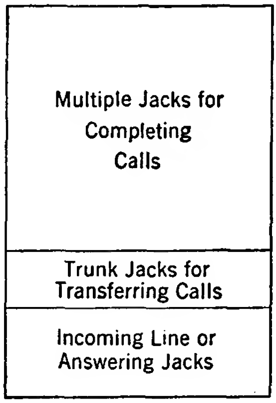


FIG. 15. Typical arrangement of jacks in a common-battery multiple switchboard.

Since one operator can reach conveniently only about 10,000 multiple jacks, a city having more than this number of telephone users must have more than one central office. When a customer connected to one central office wishes to talk with one connected to another office, the calls must be trunked to the second office for completion. Each operator must accordingly be supplied with trunks to each of the other offices in an exchange. Thus, in addition to the multiple calling jacks and answering jacks, trunk jacks appear before the operator somewhat as shown in Fig. 15.

Operation of the Common-Battery, Multiple Switchboard. In Fig. 16 is shown a simplified circuit of a typical common-battery

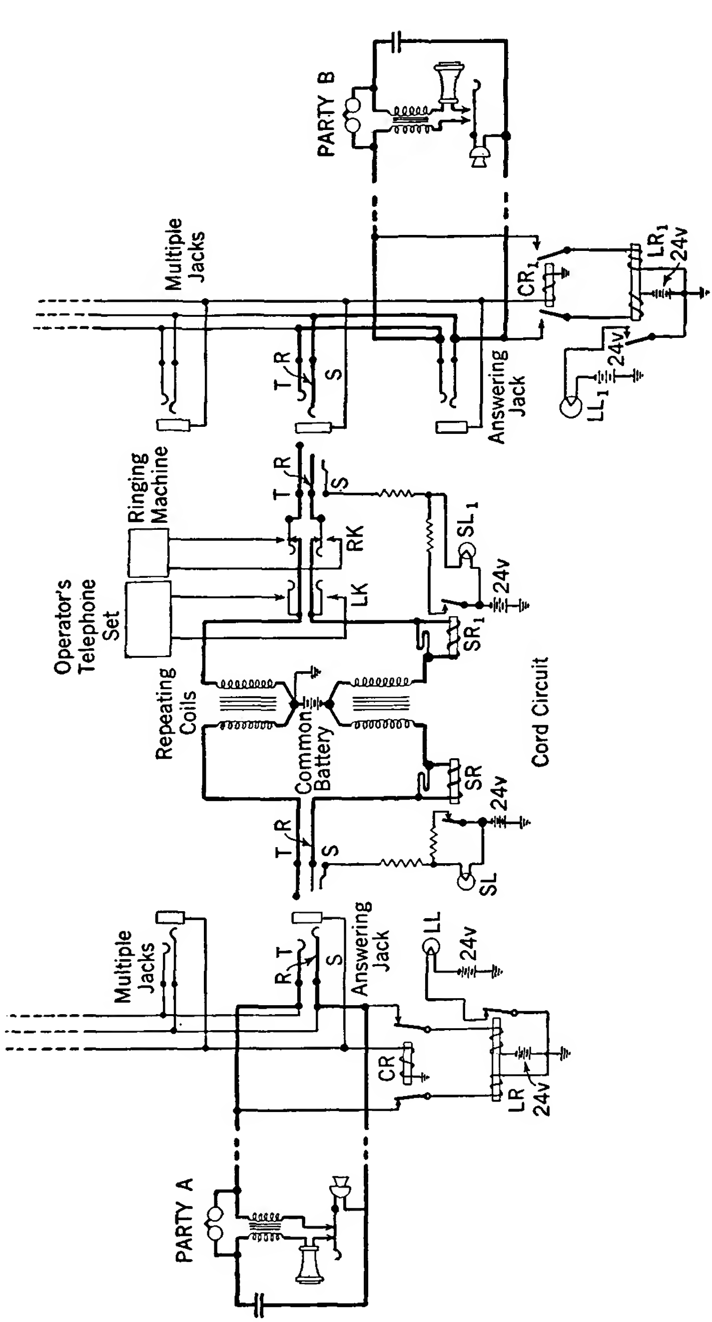


Fig. 16. Simplified circuit of a typical common-battery, multiple switchboard. Sidetone telephone sets are shown; however, antisidetone sets may be used.

multiple switchboard. The operation of this board when a call is completed from a party in one central office to a party in the same central-office district will now be considered.

To call the operator the customer at A removes the receiver from the hook, thus allowing the switch to close and permitting current to flow from the battery up through the line relay LR, through the line wires, and through the transmitter of set A. This energizes the line relay, pulling up the armature and lighting the line lamp LL in the face of the switchboard near the line answering jack. Upon observing the lighted lamp, the operator inserts the plug of an answering cord into the answering jack indicated by the lighted line lamp.

When this plug is inserted, the "tip," "ring," and "sleeve" (indicated by T, R, and S) connections are made. The cutoff relay CR is then energized over the circuit completed by the sleeve connection from the battery below the supervisory lamp SL, and the armature is pulled up, opening the line-lamp circuit. At the same time direct current flows from the central-office battery through the supervisory relay, SR, pulling up the armature of this relay, which shunts the lamp SL and prevents it from lighting. This direct current provides the transmitter talking current.

The operator's set is connected by throwing the listening key LK, and the desired number is taken by the operator. The talking currents are shunted through the non-inductive windings shown above the supervisory relays SR and thus do not have to pass through the inductive windings.

After ascertaining the desired number, the operator tests the called line by tapping the sleeve of the multiple jack of that line with the tip of the plug of the calling cord associated with the answering cord previously used. A click notifies the operator that the line is busy. If the line is not busy, the operator inserts the plug in the jack. When this is done, the tip, ring, and sleeve connections are completed, but, owing to the fact that the receiver is on the hook, no transmitter current flows to the called telephone, and thus the supervisory relay SR_1 does not operate. Current therefore flows through the supervisory lamp SL_1, through the sleeve connection and the windings of the cutoff relay CR_1, and thus the lamp SL_1 burns brightly. Also, the relay CR_1 pulls up the armatures opening the contacts in series with the line relay LR_1. When the ringing key is operated, the calling party is disconnected from the line, and the bell of the called party is rung. When the called party removes the receiver from the hook, the transmitter current flows through the relay SR_1, pulling up the contact and extinguishing the lamp SL_1 by shunting it.

The talking circuit has now been established, and conversation takes place. Each station receives direct current for the transmitter from the central-office battery through the repeating coil windings. Transfer of energy of the talking currents from one part of the circuit to the other is made possible by the repeating coils.

When either or both parties replace the receivers on the hook, the direct-current path is broken, and the two supervisory relays release, removing the shunts from the supervisory lamps, thus permitting them to light and to signal the operator to remove the connection.

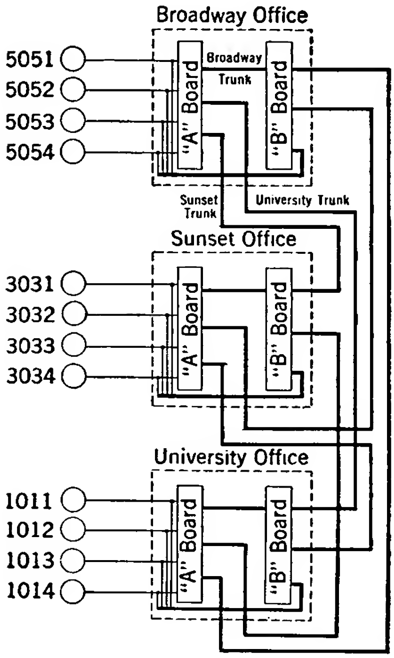


FIG. 17. Interconnections of offices in a multioffice exchange.

Telephone Service in Multioffice Exchanges. In a large exchange consisting of many central offices a small percentage of the incoming calls from the lines connected to an office is completed within that office. Many of the incoming calls are for telephones connected to other offices and must be transferred for completion over *trunks* connecting the two offices. As the number of trunk jacks is increased (Fig. 15), there is less room for the calling multiple jacks; also, this multiple is used less, and the trunk jacks more. Furthermore, in each central office additional separate multiples must be provided for completing calls trunked to that office from other offices. A point is therefore reached at which it is best to *trunk all calls before completion* even though some of the calls may be for lines connected to the office originating the call.

In the offices of multioffice exchanges it is common to have the functions of answering and completing the calls performed by two operators working at separate boards. These are known as "A" and "B" operators, and the switchboards at which they work as "A" and "B" boards. The "A" operator takes the incoming calls, and the "B" operator completes the calls. The overall working of a typical system is illustrated schematically in Fig. 17.

In considering this system, assume that the party using line 5051 connected to the Broadway office wishes to talk to Broadway 5054. The "A" operator will place a cord-circuit plug in the jack of the calling line and will take the call as explained in the preceding section. Instead of completing the connection, the "A" operator will pass the call to a "B" operator for completion. In one system that has been used, the "A" operator calls the Broadway "B" operator over a **call circuit,** and states "Broadway 5054."

The "B" operator assigns a vacant Broadway trunk to the "A" operator, who then inserts the other end of the cord which is being used in the 5051 answering line jack into this vacant trunk jack. The "B" operator then completes the call by placing the plug of the trunk assigned to the "A" operator in the calling multiple of the desired line. The called party is then rung, usually automatically.

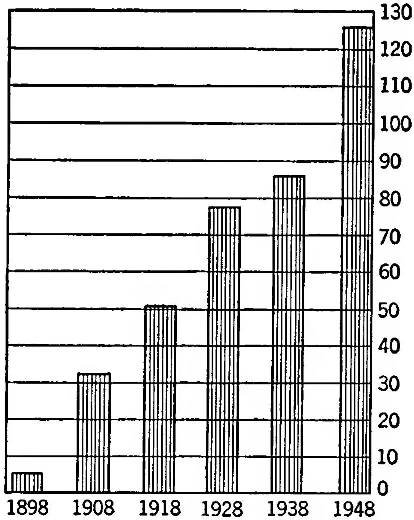


FIG. 18. Average number of daily total telephone conversations in the United States in *millions*.

If the party using the line 5051 of the Broadway office desires to talk to University 1014, the procedure is similar. When the line lamp lights, indicating that a call is to be placed, the operator answers by plugging an answering cord into the calling line jack. The operator, informed that University 1014 is desired, passes the call to a "B" operator in the University office much as explained in the preceding paragraph.

The **call-circuit method** of passing calls was considered in the preceding discussion. Often the **straightforward method** is used. With this plan of operation, the number that is desired is passed by an "A" operator to a "B" operator over the circuit that is used for completing the call.[15] The magnitude of the task of switching telephone calls is indicated by Fig. 18.

Common-Battery Connections. As explained in the discussions accompanying Figs. 14 and 16, direct current is supplied from the central-office common battery to the transmitters of the telephone sets. If the **cord circuits** of these two figures are examined, it will be

observed that the methods of connecting to the common battery are different.

The simplest method of connecting several telephone sets to a common battery is shown in Fig. 19. The inductors (choke, or retard, coils) offer low resistance to the flow of direct current to transmitters A and B. However, they present high impedance to the alternating speech currents which, therefore, flow through the telephone sets and are not short circuited by the common battery. This method is sometimes used in interphone systems for offices, etc. Often, only one inductor is used, in which event the circuit is unbalanced.

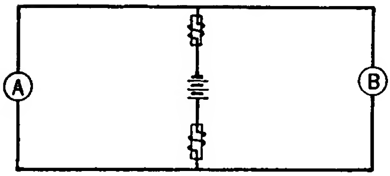


FIG. 19. One method for connecting telephone sets to a common battery. This is often used with simple telephone systems such as interphones.

The common-battery connections of Figs. 14 and 16 are shown in Fig. 20. These are more widely used, particularly the transformer, or repeating-coil, method. A simple direct-current analysis will show that Fig. 19 is a series-parallel combination and that the resistance of each circuit affects the amount of transmitter current that flows to the other. Such an arrangement is, therefore, suitable only when the resistances of the two circuits that are connected are approximately the same. If the internal resistance of the common battery is negligible, and usually it is, then with the connections of Fig. 20 the resistance of one circuit does not affect the direct current that flows to the transmitter of the other.

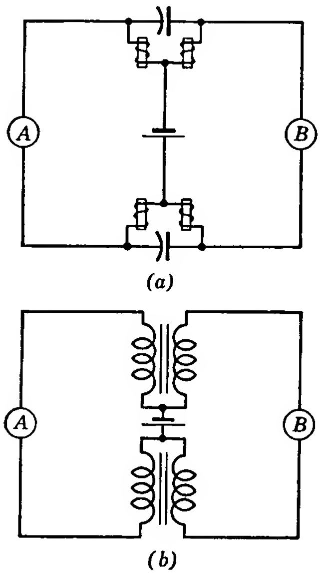


FIG. 20. Impedance-coil method (a) and repeating-coil method (b) for connecting two or more telephones to a common central-office battery.

Arrangement of Central-Office Equipment. As indicated in Fig. 21, the various subscriber lines entering a central office are in cables that terminate on the line side of the **main distributing frame,** abbreviated M.D.F. At this point, protection in the form of lightning arresters and heat coils is inserted in the circuits (page 568). The

cable terminations are marked with the number of the cable and with the cable-pair number for identification.

On the opposite side of the M.D.F. the terminations are marked with the multiple numbers, or with what might be termed the telephone number designations. These numbers correspond with the numbers on the calling multiple on the "B" board. These M.D.F.

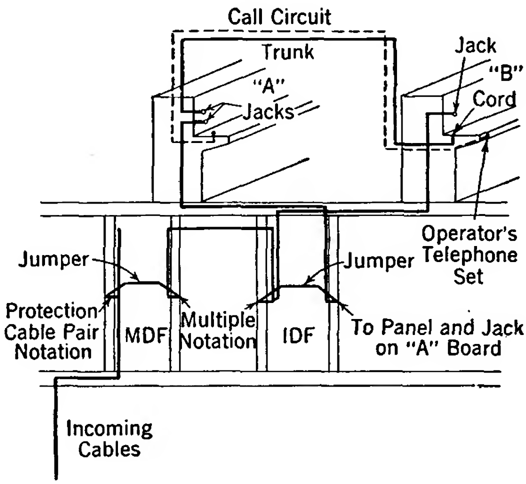


FIG. 21. Typical arrangement of apparatus for the "A"–"B" board operation with 100 per cent trunking. The "A" operator's telephone is connected to the call circuit by a key arrangement. One circuit only is shown. Switchboard cord circuits are not shown. This shows the use of a call circuit between the two operators, but often the straightforward method is used (page 367).

terminations are permanently connected with the **intermediate distributing frame** or I.D.F. terminations as shown in Fig. 21. Wires also lead from I.D.F terminations to the multiple on the "B" board. From the opposite side of the I.D.F., wires go to the answering jacks on the "A" board. These are marked with switchboard panel and jack designations.

One purpose of the M.D.F. and the I.D.F. is to provide a flexible method for terminating the telephone sets on the switchboards, and thus to provide a means for distributing the switchboard load. When a customer requests telephone service a vacant pair in the cable near his residence is selected. A vacant multiple jack on the "B" board is also chosen, and the "A" board panel and jack determined. These latter must be at a position where the "A" operator is not already fully loaded during the busy hour. As can be seen from Fig. 21, after these

selections are made, twisted wires called jumpers are used to connect the line to the corresponding points on the M.D.F. and I.D.F. That is, a jumper is placed on the M.D.F. between the cable-pair termination and the multiple termination on the opposite side of the frame. This completes the connection to the multiple on the "B" board. The multiple side of the I.D.F. is then connected by a jumper to the proper panel and jack terminations on the opposite side of the I.D.F. In this manner the connections are made to the "A" board.

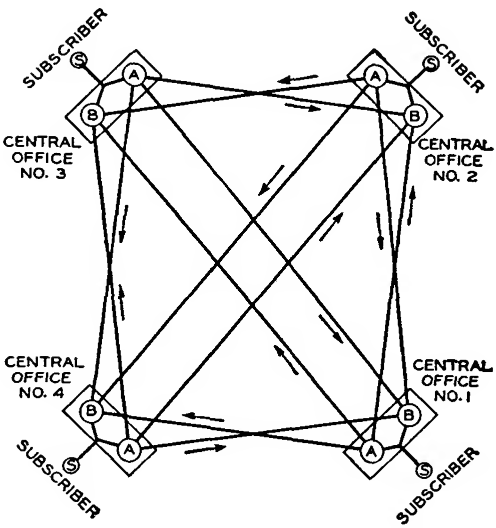


FIG. 22. Between each two offices in a multioffice area two groups of trunks are required, each group connecting the "A" board of one office with the "B" board of the other. (Reference 16.)

Tandem Trunking. In a very small community, limited telephone service could be given without a central office for connecting the lines. Each telephone user, however, would require a telephone line to each of the other subscribers. To use a simple illustration, it is easily shown[16] that without a central-office switchboard a total of 28 connecting lines and 56 line jacks (seven at each location) would be required to give service between *only eight* telephone users. With a simple switchboard, however, only eight lines and eight jacks are required.

From the comparison just given, and from the discussions of telephone trunking in multioffice exchanges (page 366), it is evident that connecting the different central offices in an exchange with trunk circuits offers a problem similar to connecting telephone sets. Just as a number of lines and jacks can be saved by using a switchboard to connect telephone sets, so a **tandem office** and a **tandem switchboard** can be used to connect telephone offices and thus reduce the number of trunks and the amount of central-office equipment required.[17] It will be recalled that, in a large exchange having many offices, a very large percentage of the calls originating in one office must be trunked to another office for completion.

The saving in trunks and equipment is indicated by Figs. 22 and 23. The tandem office may be considered a central office for switching central offices. This principle is used in both manual and dial systems.

Dial Telephone Systems.[1] In these systems the desired telephone connection between two persons is established by electrical and mechanical apparatus in **dial central offices.**[1] The operation of the apparatus is controlled by the manipulation of the dial on the telephone set of the calling party.

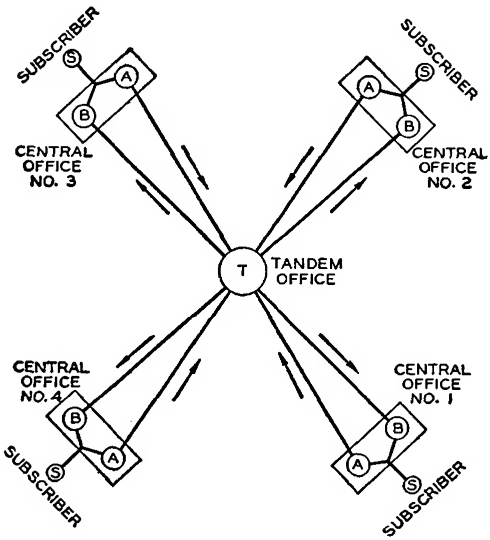


FIG. 23. When the tandem plan is used, only two groups of trunks to the tandem office are required. (Reference 16.)

When the dial telephone handset (or the telephone receiver) is removed from position, prior to dialing, the calling subscriber line is immediately connected to the central-office switching equipment. When the calling party dials the desired number, the connection is extended through various switching stages to the desired subscriber line.

If the called line is busy, busy tone is sent to the calling party. If the called line is idle, the bell of the called party is rung. When the telephone is answered, the central-office equipment completes the talking circuit and holds connection during the talking period. When both parties replace their receivers, the central-office equipment is released immediately for use on other calls.

Dial systems appear complicated and difficult to understand; however, the basic principles are relatively simple if a system is analyzed. Nevertheless, it is recognized that those entirely unfamiliar with dial systems may have to do considerable reading elsewhere to understand, in particular, the details of the panel and crossbar systems.

Five different types of dial systems are used in the United States: (1) the step-by-step or Strowger system, (2) the all-relay system, (3) the rotary dial system, (4) the panel system, and (5) the crossbar system. Each of these systems is controlled by electric impulses trans-

mitted from the dial of the calling party to the central office over the subscriber line, or loop. A brief description of the operation of the dial will now be considered before studying the different types of central-office equipment.*

The Dial.[1] A typical dial consists of the following parts: finger plate, number plate, impulse springs, impulse cam, speed-control governor, shunt springs, and driving mechanism. The dial is operated by

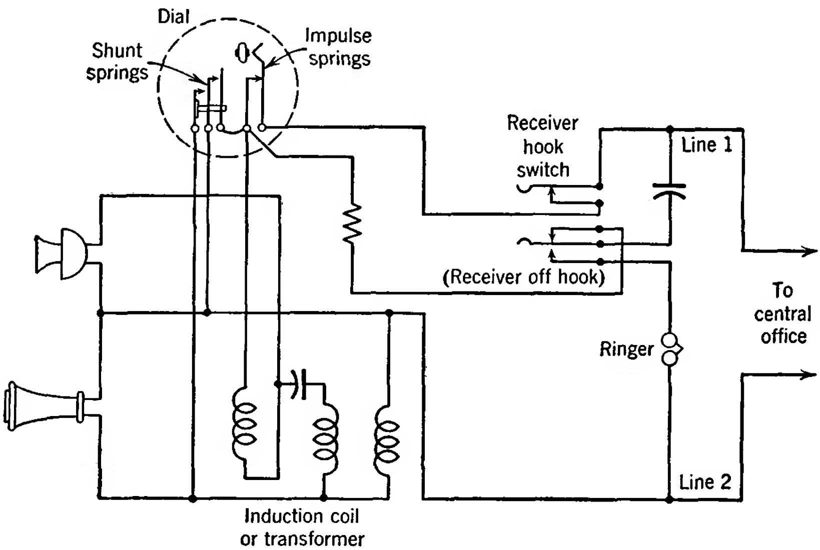


FIG. 24. Circuit of a typical common-battery antisidetone dial telephone set. The transformer windings are on a common core (not shown).

inserting the finger in the appropriate hole and turning the dial in a clockwise direction. This winds a spring with sufficient tension to return the dial to normal. When the dial is released and returns to normal, a pawl engages in a ratchet which is fastened to a shaft. This shaft is geared to a governor and to an impulse cam which opens and closes the impulse springs shown in Fig. 24 in accordance with the selected digit, and at a regulated speed of approximately 10 interruptions per second.

The shunt springs close when the dial is first rotated and remain closed until the dial returns to normal. The shunt is placed on the receiver (Fig. 24) to prevent a clicking noise during dialing, and it is

* Most of the material on dial systems in this book was prepared by Mr. Dwight L. Jones. (See Preface.)

placed on the transmitter to keep transient current impulses from flowing through it.

The basic principles of dial-operated equipment and the dial by which it is controlled have been presented. The details of the five different dial switching systems will now be considered.

Step-by-Step Dial Equipment.[4, 18, 19] Step-by-step central-office equipment consists of three fundamental units: (1) **line switches** or **line finder switches,** (2) **selector switches,** and (3) **connector switches,** all of which *may* function in establishing a connection between two lines. (In a very small exchange the selector switches are not required as will be shown later.)

When a subscriber lifts the handset or the receiver, the line switch or line finder switch connects the calling line to a selector switch (or connector switch in a small installation), thereby closing a circuit from the subscriber line to the central-office equipment. When the dial is operated, this circuit is interrupted at regular intervals (depending upon the digit dialed), and the central-office switches are operated to connect the calling line to the called line.

Because much of the step-by-step apparatus is based on a two-motion switch invented by Strowger in about 1887, this switch will be described before proceeding with the operation of the step-by-step system.

The Strowger Step-by-Step Switch.[18, 20, 21] The Strowger switch consists of three main parts: (1) the relays, (2) the bank assembly, and (3) the switching mechanism including the shaft and wipers, magnets, frame, etc., as shown in Fig. 25.

Relays. A **relay** is defined[1] as "an electromechanical device by means of which a change of current or potential in one circuit can be made to produce a change in the electrical condition of another circuit." The relays to be considered are operated by current changes.

The magnetic circuit of a relay consists of the coil core, frame or base (sometimes called "heelpiece"), and armature, all made of good magnetic materials annealed to provide a good magnetic path. The coil is wound with many turns of insulated copper wire having the desired resistance and number of turns to meet circuit requirements. In dial-telephone applications it is sometimes necessary to have relays that are slow to close or slow to release. This time delay can be produced by locating a copper slug, or sleeve, on the core. This acts as a short-circuited turn, and any change in magnetic flux linking this slug will cause current flow in the slug, which will produce magnetic flux in opposition to the flux change. A copper slug on the armature end of the coil core will make the relay slow to close and also slow to

release. A copper slug on the opposite, or heel, end will make the relay slow to release but will not affect the closing time.

The contacts of the telephone-type relays are usually of platinum, palladium, silver, or other special metals on nickel-silver or phosphor-bronze springs. The spring assemblies are actuated by the movement of the armature and provide the desired circuit switching connections.

FIG. 25. A Strowger step-by-step connector switch. (Courtesy Automatic Electric Co.)

Bank Assembly. The step-by-step switch bank assembly is made up of two banks of contacts shown at the bottom of Fig. 25. The upper of the two is called the **sleeve bank** and consists of 100 contacts arranged in 10 horizontal rows of 10 contacts each. These contacts are arranged in a semicylindrical form so that a pair of spring **wipers** mounted on the switch shaft may make contact with any of the 100 contacts.

The lower of the two is called the **line bank** and has 200 contacts also arranged in 10 horizontal rows or levels, each consisting of ten sets of two contacts, the two contacts of each set being placed one above the other with a thin insulator between them. Another pair of wipers on the switch shaft makes contact with the line bank. The two pairs of wipers are so arranged on the shaft that when the sleeve wiper is raised to a given level of the upper bank contacts, the line wiper is raised to the corresponding level on the lower, or line, bank.

Switching Mechanism. The switching mechanism of the step-by-step switch consists of the vertical, rotary, and release magnets and

associated armatures which control the motion of the shaft and wipers of Fig. 25. The vertical armature, actuated by the vertical magnet, steps the shaft up one step for each dialed impulse where it is held. The rotary magnet actuates the rotary armature to step the switch in the rotary direction where it is held at the desired position. When the release armature is operated by the release magnet, the shaft and wipers are allowed to return to the normal unoperated position.

Line Switches. The line switch is a switch connected to each subscriber line and has access to a number of trunks to succeeding switches in the step-by-step system. This line switch operates automatically when the calling party raises the handset or the receiver, and it connects the calling line to a trunk leading to an idle selector switch (or connector switch in small installations). These switches are referred to as **non-numerical switches** because they do not require dial impulses for their operation. There are two common types of line switches: (1) the **plunger type** with associated **master switch** and (2) the **rotary type.**

The plunger line switches are arranged in groups of 25 to 100 line switches under the control of one master switch and have access to ten trunks to selectors (or connectors). The plunger line switch consists of a slow-operate type of line relay and an operating magnet which actuates a pivoted plunger. This plunger may engage any one of ten sets of bank contacts connected to trunks leading to selectors. The master switch keeps the plungers of all the idle switches of the group resting opposite an idle trunk; therefore, the master switch preselects the trunk before the call is initiated.

The rotary line switch is sometimes used instead of the plunger line switch. A rotary line switch is associated with each line and is entirely independent of any other rotary line switch and, therefore, does not require a master switching mechanism. The rotary line switch consists of a set of wipers which is rotated over a semicylindrical bank of contacts by means of a magnet-operated rachet mechanism. Each rotary switch has an associated relay assembly consisting of a slow-operate type of line relay and cutoff relay.

Line Finder Switches.[22] The line finder is another type of non-numerical switch which connects the subscriber line to an idle trunk leading to a selector (or in a small system to a connector). The line finder serves the same purpose as the line switch except that *the method of accomplishing the connection is reversed.* The line finder is directly associated with a selector (or connector), instead of being associated with the line, and it *finds the line* which is originating the call. The line finder switch may be of the rotary type having access

to 25 or 50 subscriber lines, or it may be of the step-by-step type having access to 100 or 200 subscriber lines which terminate on its bank contacts.

The Selector Switch.[20] The selector switch used in the step-by-step system is of the basic type described on page 373. This switch is used for making intermediate connections between the line finder and connector switch in systems *having more than 100 lines.* A group of selectors is required for each of the digits of the call number except the last two (sometimes three). The vertical motion of the selector, controlled by the incoming impulses from the dial, raises the shaft and wipers to the desired level. Then, the selector switch automatically rotates its wipers over the contacts of that level to the first idle trunk in that group. The *rotary motion* of the selector is *not* controlled by the dial.

The principle functions of the selector may be summarized as follows: (1) to hold preceding switches operated until the control circuit is extended to the succeeding switch; (2) to connect dial tone to the calling line if it is a first selector; (3) to step to the selected level and hunt for an idle trunk in that group; (4) to clear the line of all attachments after the succeeding switch is seized; (5) to return busy tone to the subscriber if all trunks are busy on the selected level; (6) to release without interference to other trunks when the calling party replaces the handset or receiver.

The Connector Switch.[21] The connector is another switch based on the step-by-step mechanism of page 373. The connector switch is *always* used for the final unit in the step-by-step system. It differs from the selector in that it operates in response to the dial impulses for *both* the vertical and the rotary motion. The two digits of the called number which control the connector are usually the last two digits (tens and units digits).

The functions of the connector switch may be summarized as follows: (1) to hold all preceding switches operated until the connection is released; (2) to step the wipers to the desired line under control of the dial impulses; (3) to test the called line for busy and return busy tone if the called line is busy; (4) to extend the connection to the called line if it is idle; (5) to clear the called line of all attachments; (6) to guard both lines against intrusion by other calls; (7) to ring the called line and to return ring-back tone to the calling party, thereby indicating that the called line is being rung; (8) to remove ringing current from the called line when the called party answers; (9) to supply transmission battery to both lines; (10) to release when

the call is completed and to cause the release of all other switches used in the connection.

Step-by-Step Dial Systems.[18, 19, 23] In a small community dial office having 100 lines or less, the line switches (or linefinders) are used to connect the calling line to an idle connector switch, as shown

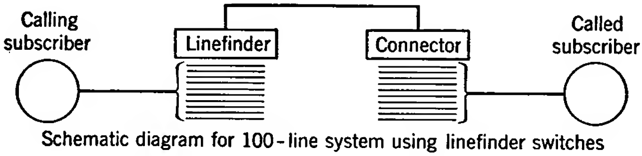


FIG. 26. The linefinder makes connection to the calling line, and the connector selects the called line.

in Fig. 26. The drawing for step-by-step switches is simplified by using 10 heavy horizontal lines to represent the 10 levels having 10 sets of contacts per level. Each single connecting line of the diagram represents several wires required for the particular circuit.

In a telephone system having more than 100 lines, a similar plan is used, except that one or more groups of selector switches are used between the line finders and the connectors. As an example, a

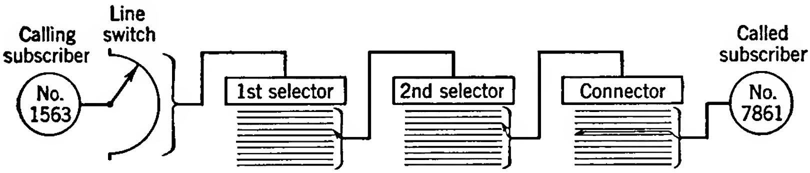


FIG. 27. A step-by-step system for a 10,000-line exchange. This uses a line switch instead of a linefinder as in Fig. 26.

10,000-line system is divided into 10 groups of 1000 lines each. Each 1000-line group is subsequently divided into 10 groups of 100 lines each. This requires two sets of selector switches as shown in Fig. 27.

Suppose calling party 1563 desires party 7861. When the receiver or handset is raised, the line switch selects a vacant *first selector*, and dial tone is sent out. The party calling 7861 then dials the digit 7. This causes the wipers of the first selector to step up to the seventh level and then to rotate until a trunk to an idle *second selector* is found. The party then dials the digit 8, and the wipers of the second selector step up to the eighth level and rotate until a trunk to an

idle *connector* is found. Dialing the third digit, 6, causes the wipers of the connector to step up to the sixth level, and dialing the last digit, 1, causes them to rotate to the first contact of that level; the desired telephone, 7861, is then rung.

Multioffice Dial Exchanges. Before considering the details of making a call in a multioffice exchange, the following additional pieces of equipment will be explained briefly.

Outgoing Trunk Switches. Outgoing trunk switches sometimes are called outgoing secondary line switches and are used as selecting switches between the selector banks and outgoing impulse repeaters.

Impulse Repeaters. Impulse repeaters are used in multioffice systems and are associated with the interoffice trunks. There are three reasons for using these impulse repeaters on interoffice trunks: *first,* they eliminate the need for the third wire (control circuit); *second,* the repeater reduces the distance over which direct current must be supplied; *third,* the repeater provides capacitors to insulate the local-office battery from the interoffice trunk.

In a multioffice step-by-step exchange having the arrangement shown in Fig. 28, the "G" office is shown in detail; however, other offices are similar, except for the manual office "S." The toll, information, and other services requiring the help of operators are concentrated at "M" office.

When a call is initiated in the "G" office, the line finder finds the calling line, connects it to an idle *local first selector,* and dial tone is received by the calling party. The first digit dialed will operate the first selector to choose the office to which the call is being made. If this digit is "G," which corresponds to the digit 4, the called line must be terminated in the "G" office, and the call will be directed to the *local second selector* in the "G" office. The second digit will operate the second selector to the level corresponding to the dialed digit, thus selecting the desired group of *local third selectors.* Likewise, the third digit will step the third selector to a level leading to the desired group of *connectors.* The last two digits operate the chosen connector switch to extend the connection to the called line.

Calls to other offices proceed in a similar manner. The first digit dialed selects the chosen office, and the outgoing trunk switch selects an idle impulse repeater and associated trunk to that office. The succeeding digits operate switches in the distant office to give the desired connection.

If it is desired to make a call to the manual "S" office, the first digit must be a seven, corresponding to the "S" on the dial. This steps the local first selector to the seventh level and selects an idle outgoing

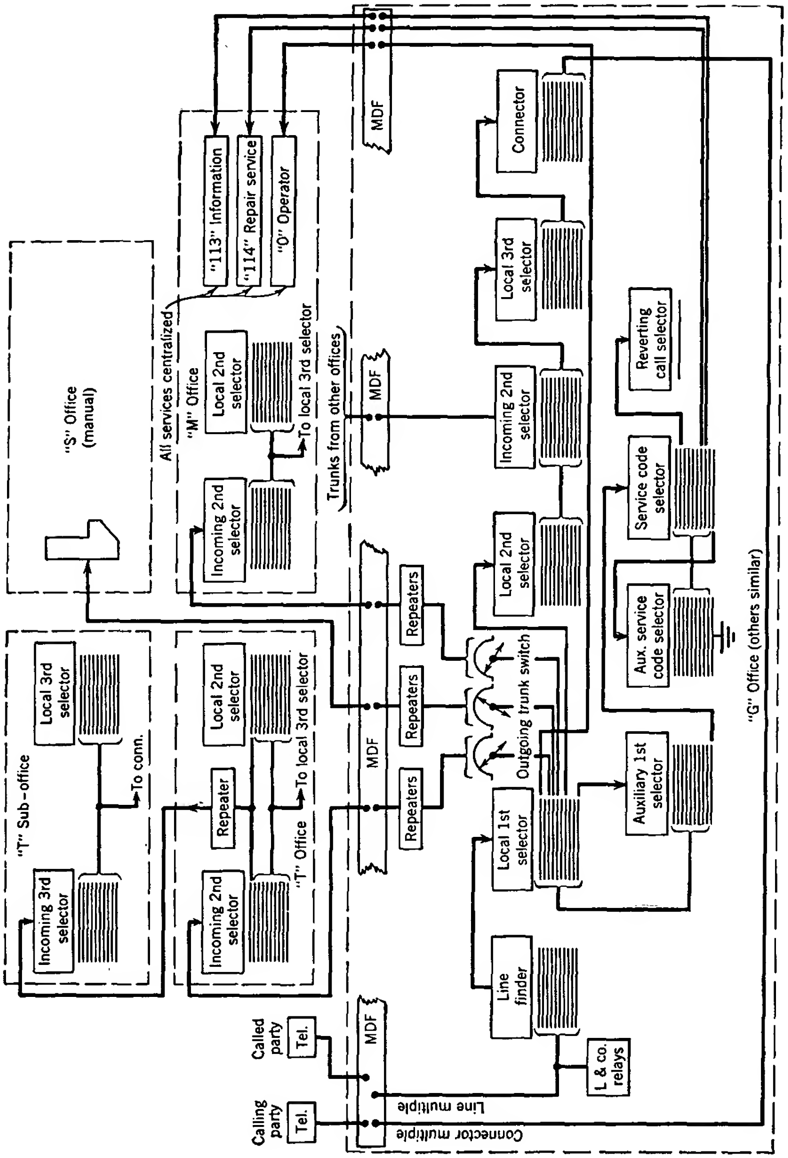


FIG. 28. Arrangement of equipment in a typical dial step-by-step multioffice exchange.

trunk switch, which in turn selects an idle trunk and associated impulse repeater to the manual office. These trunks from the dial offices usually terminate at a "B" operator's position in the manual office. The calling party may, in some systems, dial only the first digit and then speak the desired number to the operator. In other systems the complete number is dialed and the last four digits stored in **call indicator equipment** and then displayed visually in front of an idle "B" operator, who reads the number and plugs the calling line in on the proper calling multiple, thus completing the connection to the line desired.

Community Dial Exchanges. Manual switchboards in rural communities are being replaced by dial systems. These small offices, called **community dial offices,** may be left *unattended,* yet are ready to operate at all times. If the services of an operator are required by a subscriber, it is necessary only to dial "0" to obtain the toll operator at the **operator office.** This office is usually the nearest *attended* office and may handle all the operator services for several community dial offices.

A community dial office usually has a large number of party lines. These lines require either dialed code ringing or harmonic ringing to signal the desired party. Means must be provided for one party to call another party on the *same line.* A subscriber in dialing another party on the same line will ordinarily receive busy tone and the called line will not be rung. **Reverting call equipment** is used to overcome this. The calling subscriber dials either the directory number of the desired party or a special number. Busy tone is received after dialing this number; then, the calling subscriber replaces the handset, and the called station will be rung. Also the calling party's bell will be rung. When the called party answers the ringing will stop, thus indicating to the calling party that the call has been answered. The first person lifting the receiver on that line will cause the ringing to be cut off.

All-Relay Dial Systems. The all-relay system is particularly adaptable to the type of small unattended exchange previously mentioned. It is available, however, for larger systems.

Special multicontact relays that close a large number of contacts at one time are used. The one-line diagram of Fig. 29 shows the basic principles of an all-relay system.

When a call is initiated by a subscriber in the thirties group shown, the finder F-30 "tens" multicontact relay operates, connecting all 10 lines of the group through to their "units" relays. If the calling line is line 35, the finder "units" relay 5 will be operated, thus connecting

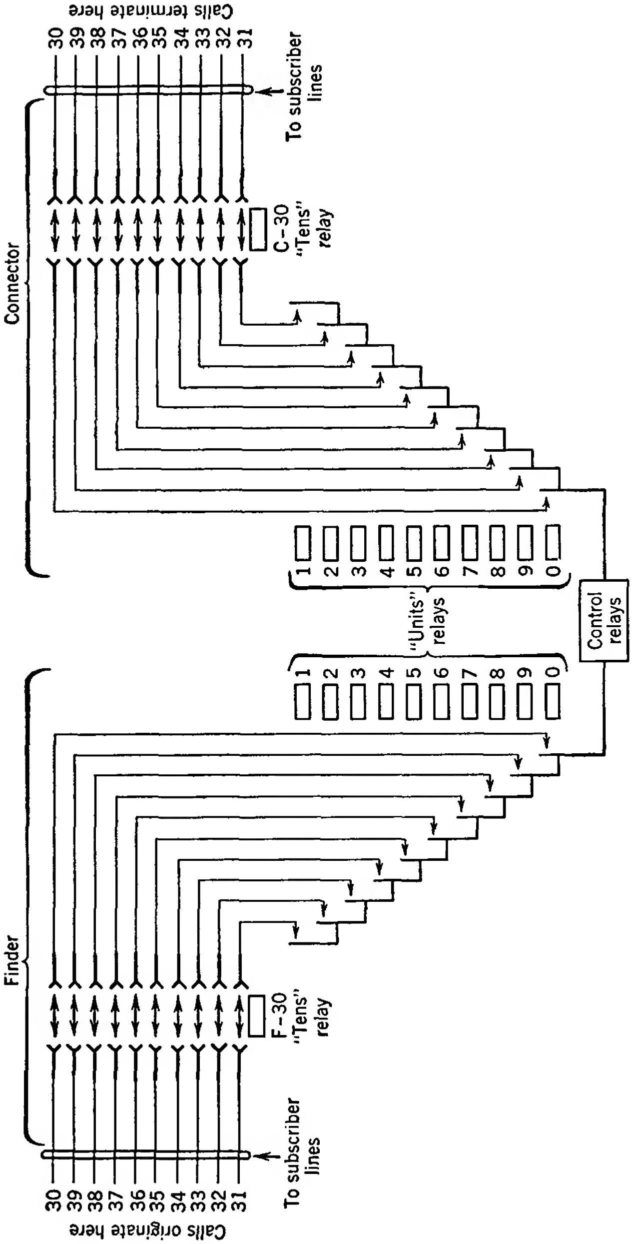


FIG. 29. An all-relay system for serving 10 lines.

line 35 to the control relays. Dialing the desired number, for instance 37, will cause the connector "tens" relay (also a multicontact relay) C-30 and the connector "units" relay 7 to operate thus connecting lines 35 and 37. A 10-line system is shown in Fig. 29; however, this basic system is extended to serve larger numbers of lines.

Rotary Dial System.[24, 25] The rotary dial system is used extensively abroad and only recently has been installed in the United States. The equipment used in the rotary dial system is characterized by the following features: *first,* the brushes of the selecting mechanisms are moved in a circular arc by a rotating member; *second,* the rotating members of the selecting mechanisms are driven by power apparatus; *third,* the dial pulses are received and stored by controlling mechanisms which govern the subsequent operations necessary in establishing the connections for a telephone call.

The switching mechanisms used in this system consist of finder switches, selector switches, sequence switches, and relays.

Rotary Finder Switch. The driving mechanism of the rotary finder switch consists of two flat gears. The driving gear is mounted on a continuously rotating shaft, and the driven gear is mounted on the center shaft of the finder switch. The driven gear is flexible and is held out of mesh with the driving gear by a downward pressure of the helical spring on the armature of the control magnet. The operation of the control magnet removes this pressure from the driven gear, which springs upward and engages the driving gear.

Rotary Selector Switch. The selector switch is similar in construction to the finder switch. The selector has two control magnets. One controls the meshing of the brush carriage driving gear, and the other controls the meshing of the driving gear to what is known as the trip spindle. The trip spindle has ten fingers which are arranged so as to unlatch one of the ten sets of brushes corresponding to the desired level or group of lines.

There are two types of selectors; *first,* the group selectors, and *second,* the final selectors.

Sequence Switch. The sequence switch is actually a power-driven relay which permits 18 different switching conditions with one complete revolution of the contacting disks. This switch is used to control ringing, busy signal, and closing of the talking circuit. The sequence switches are associated with the selectors.

References 24 and 25 are suggested for further details on the rotary dial system.

Panel Dial System.[4, 19, 26] The panel dial system was developed by the Bell System in about 1920 for use in large exchanges. The

switching in this system is accomplished on large frames on which are mounted several hundred terminals arranged in a number of groups, or banks. Connections are made to these terminals by brushes which are moved up and down on vertical rods, or elevators.

Except for the line finder, each of the major switching frames, or panels, has five of these banks. Thirty elevators may be located on each side of the frame, each having access to all the lines or trunks connected to the banks. At the bottom of the elevators are flat strips, called racks, which are controlled by magnetic clutches. When the

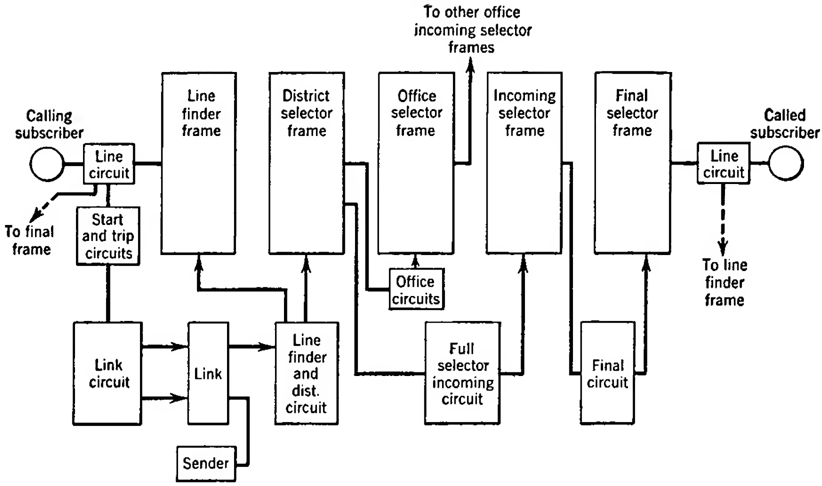


FIG. 30. Schematic diagram of the panel dial system.

elevator is to be driven upward the associated rack is pressed against continuously rotating electric motor-driven cork rolls.

When a call is initiated, an idle finder in the **line-finder frame** moves upward to connect to the calling line (see Fig. 30). The other end of the line finder terminates on a **district-selector frame.** The operation of the line finder is directed by the associated start and trip circuit, link circuit, and the line circuit of the calling line. At the same time an idle **sender** also is connected through the link circuit. The sender transmits dial tone to the calling party to indicate that the equipment is ready for the dialing. The sender proceeds to direct the completion of the call as soon as dial impulses are received. An elevator on the district-selector frame selects an idle trunk line to the proper office. In central offices of large cities, another group of frames, called **office-selector frames,** are necessary at this point to multiply

the possible number of outgoing trunks. Each incoming trunk from other central offices is terminated on an elevator of the **incoming-selector frame.** The banks of the incoming-selector frame are connected to elevators of **final-selector frames.** The incoming-selector elevator, therefore, must select a trunk to an elevator on the particular final frame to which the called line is terminated.

Crossbar Dial System.[27, 28, 29] The crossbar dial system was developed by the Bell System in about 1937 and is similar to the panel system in certain respects. However, it provides greater flexibility, and it uses a switching mechanism that is different and superior. It is probable that the crossbar system will eventually replace the panel system.

Two outstanding features of the crossbar system are the *crossbar switch,* which is used for all major switching operations, and the *marker system of control,* which is used in establishing connections. The advantages of the crossbar switch are the high speed relay-like operating characteristics and the use of a few common control circuits having the selecting and trunk-hunting features. These control circuits are used for only a few seconds in setting up a connection and then are released for use on another call. The advantage of the marker system of control is its ability to make two or more attempts to establish a path over alternate switches and trunks when the normally used paths are busy. Before it releases from a connection, it checks the circuit to ensure that the proper connection has been made. It operates an alarm when trouble conditions are encountered. The operating time of a marker is considerably less than one second, therefore, only a few markers of each type are required.

The Crossbar Switch.[29] The crossbar switch is the basic unit. This switch consists of three major parts: *first,* 20 separate vertical-circuit paths; *second,* 10 separate horizontal-circuit paths; and *third,* a mechanical means of connecting any one of the 10 horizontal-circuit paths to any one of the 20 vertical-circuit paths by the operation of electromagnets. The 20 vertical units consist of a multiple relay-like structure having 10 sets of contacts in each vertical row and a holding electromagnet. There are five horizontal bars, each provided with 20 selecting fingers (Fig. 31) made of wire. There are two electromagnets associated with each horizontal bar so that it can be operated in either of two directions.

The selecting fingers are mounted at right angles to the bar, with one finger at each vertical row of contacts. When the selecting bar is rotated through a small arc by one of its electromagnets, the selecting fingers move up or down into a position such that, if one of the holding

bars (Fig. 31) is operated by its electromagnet, the selecting finger at the crosspoint of the two bars will operate the set of contacts at that point. The selecting bar can then be released, and the selecting fingers not being used will return to normal, but the one finger used to operate the contacts will remain latched under control of the holding bar.

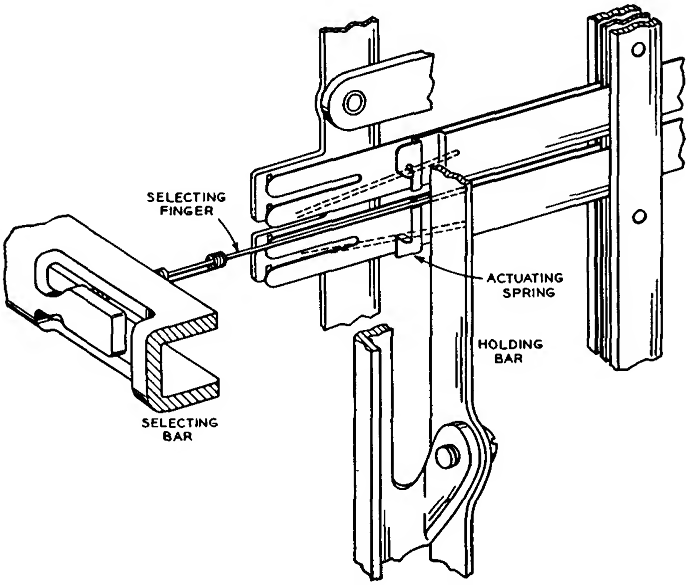


FIG. 31. Simplified schematic of the selection elements of the crossbar switch. (Reference 29.)

A single crossbar switch (Fig. 32) having 200 points with 20 vertical units, each connected to a subscriber line, and 10 trunks wired to the horizontal units, provides an arrangement for any one of the 20 lines to be connected to any of the 10 trunks. The addition of other switches, which have their vertical units wired to other groups of 20 lines and connect the horizontal units in multiple (parallel) to the horizontal units of the first switch, will provide additional lines having access to the same 10 trunks. Still greater trunking access may be obtained by using two groups of switches, referred to as "primary" and "secondary" in a link frame.

A **link frame** (Fig. 32) is made up of 20 crossbar switches arranged in two vertical files of 10 primary and 10 secondary switches. Each primary switch has 20 lines connected to the vertical units, and 20 trunks are connected to the vertical units of each secondary switch.

The horizontal multiple of each primary switch is connected (Fig. 32) to the horizontal multiple of the secondary switches so that each primary switch has one horizontal path connected to each of the 10 secondary switches. This gives any of the 200 lines access to any of the 200 trunks on the secondary switches.

Operation of the Crossbar System. In the crossbar system the selection of the talking path is under the control of three common-controller circuits. The switching frames are as follows: **line-link**

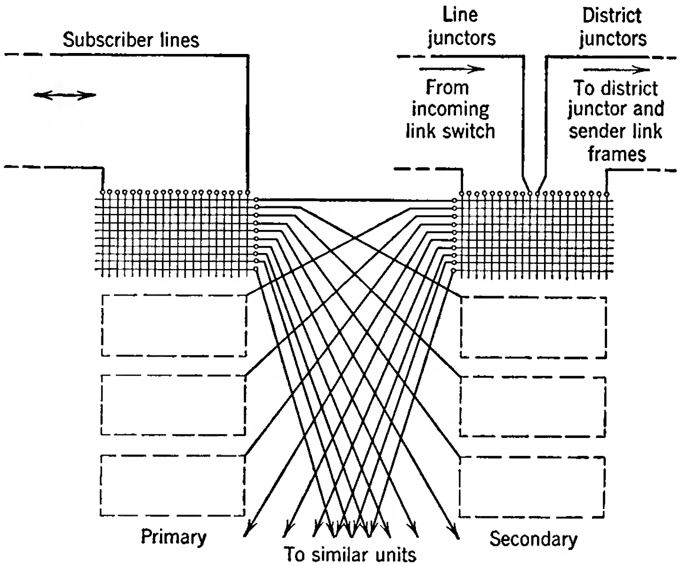


Fig. 32. A portion of the trunking arrangement of a line-link frame. Ten crossbar switches are in each vertical row, as indicated by the wording "to similar units."

frame, district-link frame, office-link frame, and **incoming-link frame.** The district, office, and incoming frames of the crossbar office are used for the same switching functions as the district, office, and incoming selector frames of the panel office, but the line-link frame may be considered as two frames since it does the work of two frames. On outgoing calls it acts like a line finder connecting the calling line to the district-link frame; on incoming calls it connects the trunks from the incoming frames to the called line.

Each crossbar frame consists of primary and secondary switches. The connections between these two sets of switches in the same frame are called "links," and the connections between frames are called "junctors."

The progress of a call through a crossbar system can be divided into the four following stages.

First Stage. When the calling party lifts the handset or the receiver, the line relay operates and seizes a line-group controller circuit that is common to the line-link frame associated with the calling line (Fig. 33). This control circuit finds the calling line and simultaneously selects a path to an idle district junctor. Before an idle line link to the district junctor is chosen, the control circuit determines which of the district junctors are available to the line links and also determines which of these has access to an idle sender. The primary selection is then made, connecting the line link through to the district junctor. This junctor is connected to a particular sender-link frame, and the associated sender-link control circuit then selects an idle subscriber. After this connection is completed between the calling line and the originating or subscriber sender, the two control circuits are released and made available for other calls. The originating subscriber sender furnishes dial tone to the calling party, indicating that dialing may be started.

Second Stage. The originating, or subscriber, sender connects to the originating marker through its connector circuit, and, as the number is dialed, the sender transfers the called-office code and the district-link frame identification to the marker circuit. Through the identification of the district junctor that has been chosen, the marker knows the incoming end-point for this stage of the connection. Also, by selecting an idle trunk leading to the called office, the marker can establish the outgoing end-point for this stage of the connection. Now the marker selects an idle path between these two points to complete this stage of the connection. This path will consist of three sections: the district link, office junctor, and office link. Some of these possible paths may already be in use on calls of other combinations; therefore the marker tests each link of a path before choosing the route. When the marker has completed the connection between the district junctor and the outgoing trunk, it checks the connection to ensure that it has been properly made and then releases itself from the connection. The connection is then held under the control of the district junctor. The marker completes its functions in approximately 0.5 second and then is released for use on other calls.

When the district junctor has been connected to the outgoing trunk, the originating sender closes a bridge across the trunk conductors, thus operating the line relay of the incoming trunk circuit in the terminating office.

Third Stage. The sender-link control circuit (Fig. 34) associated with the sender-link frame on which the incoming trunk appears establishes a connection between the incoming trunk and the terminating sender. This sender-link control circuit is then released for other calls. At this time the calling subscriber is still connected to the originating sender, and the dialing may not yet be completed. When the dialing is completed the originating sender transfers the dialed number to the terminating sender and then releases itself from the connection. The calling line is now connected through the district junctor to the incoming trunk.

Fourth Stage. The terminating sender now has the dialed number, connects to the terminating marker through the terminating-marker control-circuit, and transfers the called line number and the incoming-link identification to the marker. The marker then sets up a preliminary connection to the called line

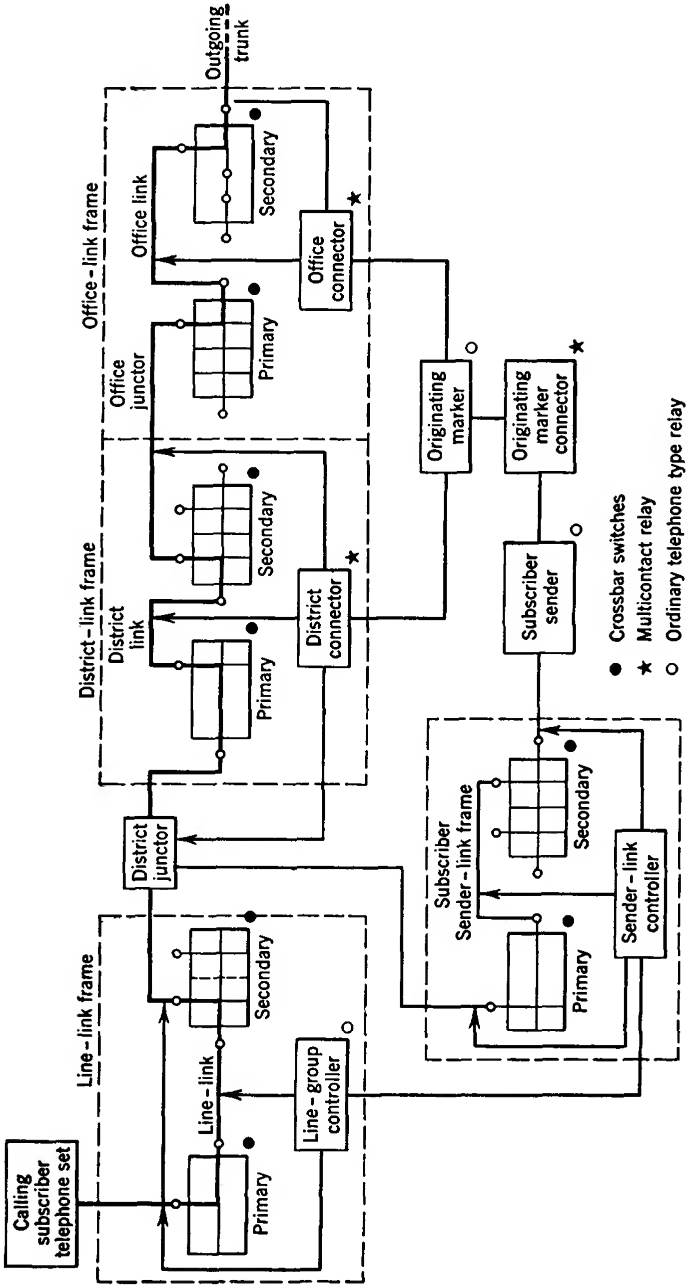


FIG. 33. Simplified schematic diagram of originating equipment in the crossbar dial system.

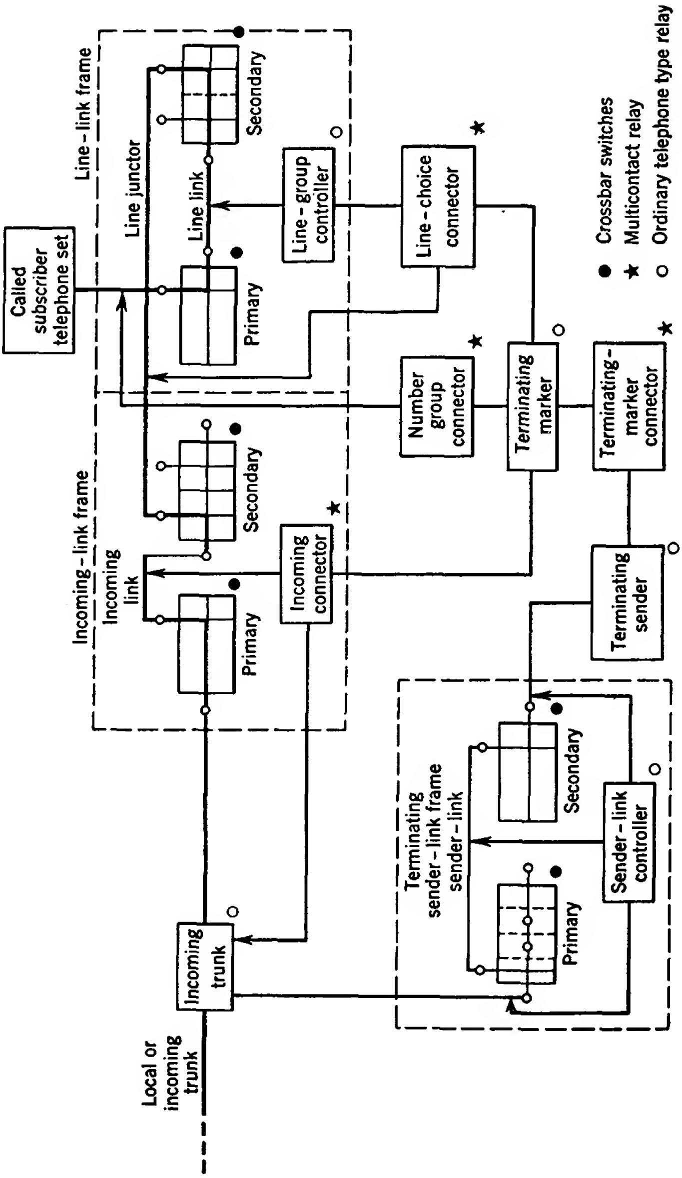


FIG. 34. Simplified schematic diagram of terminating equipment in the crossbar dial system.

through the associated "number-group connector" and a frame on which the called line appears in numerical sequence. The marker tests for a busy condition.

If the called line is idle, the terminating marker will connect through the line-choice connector to the line-group control circuit. The marker must then select an idle path through the incoming-link frame and the line-link frame in a manner similar to that used in the district-link and office-link frames. This path will consist of three parts: an incoming link, line junctor, and line link. After the connection is made to the called line, the marker causes the incoming trunk circuit to start the ringing of the called line and return ring-back tone to the calling line. The terminating marker, sender, and sender-link frame are then released, and the transmission circuit between the calling and called line will be completed when the called party answers.

When the terminating marker finds the called line busy, it will cause the incoming trunk circuit to return busy tone to the calling line.

Crossbar Senders. The functions of the crossbar originating senders thus far explained have been based on the assumption that the call was to a crossbar dial office. These sender circuits are, however, arranged for all possible conditions, such as for calls to a panel or manual office. In such instances the sender guides the panel selectors in completing a call to a panel office or operates a call indicator (page 380) in a manual office.

REFERENCES

1. *American Standard Definitions of Electrical Terms.* A.I.E.E., 1941.
2. Johnson, K. S. *Transmission Circuits for Telephonic Communication.* D. Van Nostrand Co.
3. Rhodes, F. L. *Beginnings of Telephony.* Harper & Bros.
4. Kloeffler, R. G. *Telephone Communication Systems.* The Macmillan Co.
5. Shepardson, G. D. *Telephone Apparatus.* D. Appleton & Co.
6. Gibbon, C. O. An Explanation of the Common Battery Anti-Sidetone Subscriber Set. *Bell System Tech. J.*, April, 1938, Vol. 17, No. 2.
7. Pye, H. C. Transmission Characteristics of Types 40 and 50 Monophones. *Strowger Technical Journal,* December, 1939, Vol. 8, No. 2.
8. Inglis, A. H. Transmission Features of the New Telephone Sets. *Elec. Eng.*, October, 1938, Vol. 57, No. 10.
9. Jones, W. C. Instruments for the New Telephone Sets. *Elec. Eng.*, October, 1938, Vol. 57, No. 10.
10. *X-ray Story of the Kellogg Masterphone.* Kellogg Switchboard and Supply Co.
11. Colpitts, E. H. Scientific Research Applied to the Telephone Transmitter and Receiver. *Bell System Tech. J.*, July, 1937, Vol. 16, No. 3.
12. Ingram, S. B. The 313A Vacuum Tube. *Bell Lab. Record,* December, 1936, Vol. 15, No. 4.
13. Ingram, S. B. Cold-cathode Gas-filled Tubes as Circuit Elements. *Elec. Eng.*, July, 1939, Vol. 58, No. 7.
14. *Bell Lab. Record,* January, 1943, Vol. 21, No. 5, p. vi.
15. Gherardi, B., and F. B. Jewett. Telephone Communication Systems of the United States. *Bell System Tech. J.*, January, 1930, Vol. 9, No. 1.

16. SPENCER, C. G. The Manual Tandem Board. *Bell Lab. Record,* January, 1930, Vol. 8, No. 5.
17. BRONSON, F. M. Tandem Operation in the Bell System. *Bell System Tech. J.,* July, 1936, Vol. 15, No. 3.
18. SMITH, A. B., and W. L. CAMPBELL. *Automatic Telephony.* McGraw-Hill Book Co.
19. CRAFT, E. B., L. F. MOREHOUSE, and H. P. CHARLESWORTH. Machine Switching Telephone System for Large Metropolitan Areas. *Bell System Tech. J.,* April, 1923, Vol. 2, No. 2.
20. The Selector. *Bulletin* 808, Automatic Electric Co.
21. The Connector. *Bulletin* 807, Automatic Electric Co.
22. DODGE, W. L. Development of Step-by-step Line Finders. *Bell Lab. Record,* February, 1929, Vol. 7, No. 6.
23. BUTZ, E. D. An Outline of Step-by-step Operation. *Bell Lab. Record,* December, 1929, Vol. 8, No. 4.
24. HATTON, W. Operating Principles of the 7-A2 Rotary System. *Elec. Commun.,* September, 1936, Vol. 23, No. 3.
25. HATTON, W. Rotary Automatic Telephone System. *Elec. Eng.,* July, 1948, Vol. 67, No. 7.
26. COLLIS, R. E. The Panel System. *Bell Lab. Record,* July, 1931, Vol. 9, No. 11.
27. SCUDDER, F. J., and J. N. REYNOLDS. Crossbar Dial Telephone Switching System. *Elec. Eng.,* May, 1939, Vol. 58, No. 5.
28. Crossbar. A series of articles reprinted from *Bell Lab. Record,* Vol. 17, 1939.
29. REYNOLDS, J. N. The Crossbar Switch. *Bell Lab. Record,* July, 1937, Vol. 15, No. 11.

REVIEW QUESTIONS

1. Distinguish between the terms central office and telephone exchange.
2. Distinguish between the terms telephone set and telephone station.
3. From the standpoint of the location of the battery, how may telephone systems be classified?
4. From the standpoint of the method of switching, how may telephone systems be classified?
5. Why is a capacitor sometimes used as indicated in Fig. 3?
6. What is the difference between a sidetone-reduction circuit and an antisidetone circuit?
7. What two principles are used in antisidetone circuits?
8. Discuss the methods used for making line connections in common-battery systems.
9. Describe a typical subscriber line. About what would be the resistance of a typical subscriber line, or loop?
10. In a common battery system, when a "B" operator starts to plug in on a called line, how does she know if the line is busy?
11. What is meant by a cord circuit, and what equipment does it contain?
12. Explain the principle of tandem trunking.
13. What types of dial systems are used in the United States?
14. Discuss the functions of the shunt springs in the dial.
15. About what is the normal speed of a telephone dial? Why not speed it up, so that dialing can be faster?

16. What is an important difference in the operation of a selector switch and a connector switch?
17. In the step-by-step system, what devices are used to interconnect the central-office equipment and the calling subscriber line?
18. What is the bank assembly in a step-by-step system? A wiper?
19. What major functions are performed by a selector switch? By a connector switch?
20. Name several uses for step-by-step connectors outside of telephone systems.
21. What is accomplished by reverting call equipment?
22. What is a typical field of application of an all-relay system? What major pieces of equipment are used?
23. What is a fundamental difference between the step-by-step system and the panel and rotary systems?
24. What is a fundamental difference between the step-by-step system and the crossbar system?
25. What types of dial systems involve the extensive use of sliding contacts to switch the circuits? What type of contact is used in the crossbar system?
26. Describe the operation of a crossbar switch.
27. What switching frames are used in a crossbar system?
28. What types of switches are on each crossbar frame?
29. What is the difference between a link and a junctor?
30. Enumerate the stages involved in the progress of a call through a crossbar system.

PROBLEMS

1. Two simple telephone sets each composed of a 50-ohm transmitter and a 75-ohm receiver are to be connected by a 500-foot 22-gauge cable circuit and fed by a 12-volt battery in series midway between the two sets. For maximum direct current through the transmitter, should the transmitter and receiver at each end be in series or parallel?
2. For what distance is the efficiency of the two connections the same as judged by the transmitter current?
3. The internal impedance as measured between the output terminals of a typical local-battery telephone set at a frequency of 1000 cycles is 433 ohms effective resistance and 0.043 henry inductance. This set delivers an alternating voltage of 0.70 volt to a 600-ohm load when the transmitter is excited by a 1000-cycle tone. Calculate the open-circuit voltage of the set.
4. Referring to Fig. 16, the direct-current resistance of the various units is as follows: each repeating coil winding, 22.5 ohms; supervisory relay (S.R.), 9.4 ohms; line relay (L.R.), 1000 ohms for each winding; transmitter, 50 ohms; receiver, 80 ohms; induction coil primary, 9 ohms, and secondary, 14 ohms; ringer, 1400 ohms. The set is connected to the central office by one mile of 22-gauge cable, and the battery voltage is 24 volts. Calculate the current which will flow from the central office, and the current through the transmitter, when the receiver is removed from the hook.
5. Make the same calculations for the circuit when the operator plugs in on the line and the receiver is off the hook.
6. The inductors or telephone "retard coils" of Fig. 19 have a direct-current resistance of 20 ohms, line A has a resistance of 600 ohms, and line B has a resistance of 100 ohms. The battery voltage is 24 volts. Calculate the cur-

rents that will flow through 100-ohm transmitters in the sets A and B. Now assume that the connections are as in Fig. 20(*b*) and that each winding has a resistance of 20 ohms. Calculate the currents that will flow through the same transmitters at A and B as before.

7. Seven central offices exist in an exchange area. Each office must have a trunk cable to each other office. How many trunks will be required for the arrangement of Fig. 22? If six offices are located on the circumference of a circle having a radius of 6 miles and one office is at the center, estimate the trunk miles saved if the tandem system of Fig. 23 is used. Assume that the number of trunk circuits between each office must be the same.
8. Draw a diagram similar to Fig. 27 for a system that theoretically will handle a maximum of one million lines. Explain how a call is completed to a given party.

CHAPTER 11

TELEPHONE TOLL SERVICE AND SYSTEMS

Introduction. Toll and long distance service, to be considered in this chapter, are not precisely the same although the terms are often used interchangeably. **Toll service** is telephone service between two parties located in different exchanges and for which a special toll charge is made. If the two exchanges are far apart, **long distance service** is the term applied. Occasionally the term **intercity service** is used.

Toll service is provided over **toll lines,** defined[1] as "a telephone line or channel between two central offices in different exchanges." A **telephone channel** is defined[1] as "a path suitable for the transmission of voice-controlled electric waves between two stations." The telephone channel providing toll telephone service may be supplied over lines, cables, or by radio.

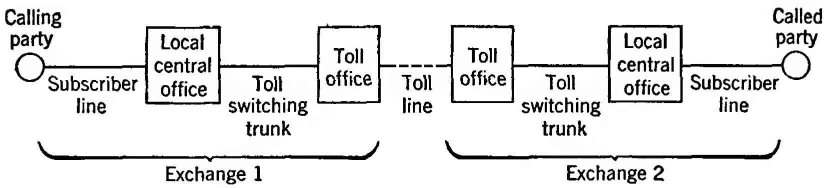


FIG. 1. Illustrating how a toll call originating in exchange 1 is completed to distant exchange 2 when a direct toll line exists between the two exchanges.

Toll lines, or more correctly, toll channels, are interconnected, or switched, at **toll offices.**[1] At present (1949) most of the switching is by **toll operators,** and manually operated **toll switchboards.** Toll dial-switching equipment is used in some instances, and this method is being extended rapidly.

Typical Toll Connection. Toll systems are designed to meet special situations and are less standardized than local systems. The basic arrangement is shown in Fig. 1.

The line of the calling party is connected by the local operator (or dial equipment) to the toll office and is there connected to a toll line,

or channel, to the distant exchange. There the line, or channel, terminates at the toll office, and connections are made through the proper local central office to the called party's line.

Toll Trunking Methods. A toll operator in one toll office contacts an operator in another toll office and passes a call over a toll line, or channel, called a **toll trunk.** The manner in which the trunks are *arranged* and *used* is known as the **toll trunking method,** or **trunking method.**[2] Several methods have been used, as follows:

Call-Circuit Trunking Method.[2] When a toll call is to be passed, the toll operator at the originating office presses a key that connects her telephone set to a **call circuit** leading to a toll operator in the desired distant toll office. The distant toll operator assigns a vacant toll channel, or trunk, to the originating operator, and the toll call is completed over this assigned trunk. This procedure has been largely replaced by the method that follows.

Straightforward Trunking Method.[2] The toll operator at the originating office connects her telephone set to a vacant toll trunk to the desired office. At this office the operator passing the call is connected automatically to an operator who will complete the call. The required information is passed, and the toll connection established, over this same circuit. Note that in the call-circuit method the trunk selection is made by the operator at the terminating office, whereas in the straightforward method the trunk is selected by the originating operator.

Ringdown Trunking Method.[2] This is used where it is not economical to install the equipment for the straightforward method. In this system the originating operator plugs in on a trunk to the desired city and presses a key which operates a signal at the distant toll switchboard. An operator connects her telephone set to the trunk and answers. The call is then passed to the distant operator and completed by her over the same trunk.

Dial Trunking Method.[2] In some installations, the operator in the originating office connects to a trunk to the called locality and dials the called number directly.

Toll Operating Methods. By toll operating methods is meant the process by which a toll call is *received, recorded, completed,* and *timed;* this process is designated[2] as "handling the call." The actual methods of handling toll calls depend, among other factors, on the toll trunking system used. In the past, a single-ticket method, a two-number method, a two-ticket method, the A-board toll method, and the combined line and recording method have been used.[2] At present, only the last two are in use, and the discussion will be limited to them.

A-Board Toll (AB) Method.[2] This is the simplest and most direct method of handling toll calls. The calling party signals his local A-board operator by lifting the handset or receiver, or by dialing "0" in dial systems. This operator takes the call, records it, and trunks it, usually by the straightforward or the dial method, to the desired point. These additional duties decrease the number of local calls an operator can handle but give the customer substantially the same service as for local connections.

Combined Line and Recording (C.L.R.) Method.[2] This is an outgrowth of the single-ticket method[2] formerly used. The calling party asks the A-board operator for, or dials, "Long Distance." The **outward toll operator** takes the call, plugs in on a trunk to the desired city, and rings on the line if ringdown trunking is being used. While waiting for the called operator to answer, the calling operator records the call. When an **inward toll operator** in the desired city answers, the call is passed and completed over the trunk being used.

These paragraphs have described the simplest calls only; person-to-person calls, for example, often introduce delays and variations of the methods considered. Also, direct circuits are not provided for all possible originating toll calls. But, in most instances, the calling party remains at the telephone until the call is completed.

Toll Switching Plans. The extent of the telephone system in the United States is indicated approximately by Fig. 2. It is evident that plans for switching toll lines must be made carefully. About 1930 a toll switching plan[3] was devised as follows:

Within a given area, such as a state, so-called **primary outlets** were chosen at important centers, and each of these outlets was *directly connected* with toll lines. Each outlying toll center in an area was directly connected to *at least one* of these primary outlets. Thus, any two toll centers in a chosen area, even though remotely located, could be connected by a maximum of only two intermediate switching operations, and in many cases with only one.

In the United States and eastern Canada approximately 150 primary outlets were required to serve the 2500 toll centers, and of these 150 outlets, eight were chosen as **regional centers.** Each primary outlet was directly connected with *at least one* of these regional centers, and each regional center was directly connected with *each of the other regional centers.* Thus, each primary outlet could be connected to any other primary outlet in the country by a maximum of only two switching operations, and within an area served by a regional center, by only one switching operation.

The original toll switching plan has been modified.[2] With the ad-

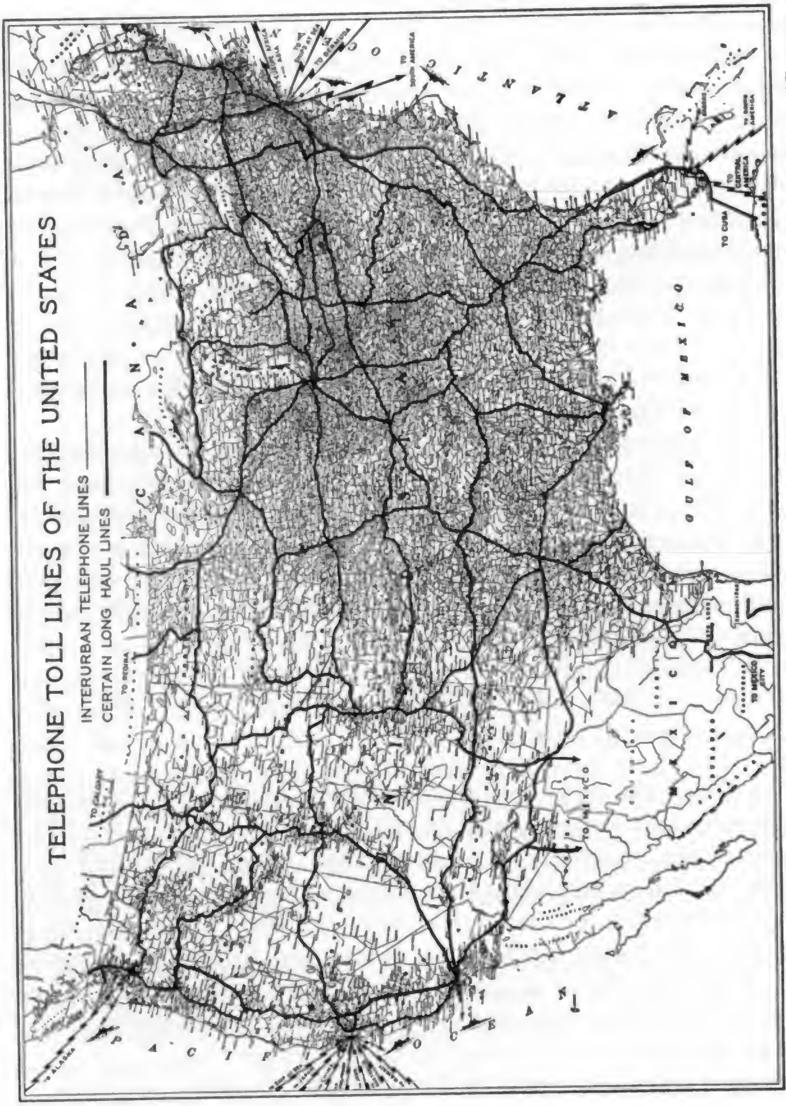


FIG. 2. Excluding certain recent minor changes, the telephone network is approximately as shown. (Courtesy Bell System.)

vent of toll dialing and a reduction of manual toll operating, further revision of the basic system of handling toll calls will occur.

Toll Switching Equipment. Special electromechanical dial and key-operated equipment have been developed for completing toll calls to subscribers connected to dial offices, and for interconnecting toll offices. The many details involved in this special equipment preclude its consideration. Suffice it to say that step-by-step and crossbar apparatus is used for these purposes.[4, 5]

Nationwide Dialing.* During the past 25 years there has been a conversion from manual to dial local offices and now there is a growing trend toward conversion from manual to dial switching for toll calls. The arrangements now being made provide that the toll operator at the calling end dial the called subscriber; however, these plans are being made so that direct dialing by the calling subscriber may be added at a later date. This will necessitate the use of **automatic toll ticketing,**[7, 8, 9, 10] that is, a means of recording the calling and called line and timing of the call for billing purposes.

One of the first problems of nationwide dialing is the selection of a numbering scheme whereby the operator can reach any particular subscriber without conflict because local numbering of subscribers in different towns is the same.[11] Before investigating a possible numbering scheme, it is necessary to examine the numbering on the dial itself. The following numbers and letters are found in the ten holes on the standard telephone dial.

Letters—		ABC	DEF	GHI	JKL	MNO	PRS	TUV	WXY	Z
Numerals—	1	2	3	4	5	6	7	8	9	0

From this it will be noticed that the letters *used in office names* appear in only eight of the holes on the dial. The first position is avoided because of the possibility of a single digit being transmitted by a "fumble" of the switch hook, thus giving a wrong number. The "0" position is reserved for dialing the operator with a single pull of the dial. For toll dialing purposes, it is proposed to set up numbering areas on a basis similar to that used for choosing office codes in local systems, but in such a way that it will not interfere with the method of dialing local calls.

In New York, for example, it was necessary to assign numerals instead of letters to the third digit,[11] therefore using only the first two letters of the office code such as ADams-2, or BEacon-3, or even

* Material for this section was prepared by Mr. Dwight L. Jones. (See preface.)

ADams-2 and ADams-3. Under this method of assigning office codes, it is possible theoretically to have $8 \times 8 \times 10 = 640$ local offices in a single numbering code area, and, with each office having 10,000 lines, the capacity of each numbering code area would be 6,400,000 subscribers. This is only theoretical, however, and actually the number per area will be considerably fewer than this figure because many offices have far less than 10,000 lines and because there are only approximately 500 office codes that are useful. It has been estimated that between 50 and 75 numbering code areas will be sufficient to serve the United States and Canada and that this would handle many times the present total of about 40,000,000 telephones in these two countries.

For dialing purposes, the United States may be divided by states in some regions and by groups of two or three states in others, and the more thickly populated states may be divided into two or three numbering code areas.[11]

Each area will be handled on a seven-digit basis, and two digits will be required for the toll code, giving a total of nine digits which will be sufficient to give every telephone subscriber a toll number without interfering with the local numbering. For calls within the same numbering code area, seven digits will suffice, but on calls to other areas it is desired for routing purposes to have something to distinguish this call from a local call. As already pointed out, neither the "1" nor the "0" appear in the first two digits of a local office code, therefore if a "1" is inserted as the second digit, the code for a call to area "65" would be "615." This makes ten the maximum number of digits required for nationwide dialing.

In reference 11, which considers the preceding matters in detail, nationwide dialing with step-by-step equipment and with crossbar equipment are compared. It is shown that with the use of crossbar toll-switching equipment nationwide dialing, using the ten-digit code, is a definite possibility. It will, no doubt, be some years before this plan is in effect, although at present (1949) the groundwork has been well laid.

Telephone Repeaters.[12] The attenuation loss of the largest size telephone open-wire line is 0.03 decibel per mile (page 222), and of the largest and heaviest loaded telephone cable about 0.27 decibel per mile (page 253). The loss that can be tolerated in a given section of line depends on many factors, such as the grade of service and the amount of crosstalk and noise present (Chapter 14). For open-wire lines the line transmission loss should not exceed about 10 decibels, and for a cable (which is less susceptible to noise) about 20 decibels. This means that for **voice-frequency telephony,**[1] effective trans-

mission over open-wire copper lines of the largest size wire requires amplifiers at most about 300 miles apart. For voice-frequency transmission over the largest and heaviest loaded cables amplifiers must be at most about 60 miles apart.

The **amplifiers**[1] used in telephone lines and cables are often small class A power amplifiers (page 291) for the **voice-frequency circuits.**[1] The entire amplifying device is called a **telephone repeater,** defined[1] as "a combination of one or more amplifiers together with their associated equipment for use in a telephone circuit."

If the signal magnitude is permitted to fall until it is comparable to the noise and crosstalk (that is, if the **signal-to-noise ratio** becomes low), then the signal is masked out and cannot be recovered. Also, if the signal is amplified at any point until it is strong compared with the signal strength in adjacent circuits, the strong signal may cause excessive crosstalk to these adjacent circuits.

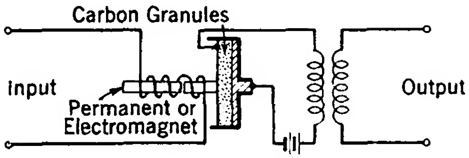


FIG. 3. Principle of the Shreeve mechanical repeater. The permanent magnet exerts a constant pull, preventing double-frequency effects (page 119).

Early Telephone Repeaters. The search for a satisfactory telephone repeater began early in the history of telephony and followed many lines.[12, 13] With the exception of the device now to be considered these investigations were without success.

As explained on page 94, the carbon-granule transmitter is an amplifier. It follows therefore that a receiver element mechanically connected to a transmitter provides an amplifying element as illustrated in Fig. 4.

The **mechanical repeater element** perfected by Shreeve was the most satisfactory device of the receiver-transmitter type, and in 1904 it was successfully used commercially between New York and Chicago.[12] As can be seen from Fig. 3, the element consists of receiving windings which actuate the movable magnetic core. This is connected to a plunger varying the resistance of the carbon-granule path, thus controlling the output from a connected battery.

The amplifying element of one model consisted of two main parts, the *cartridge* and the *socket.* The cartridge contained all the working parts likely to become defective in service and could readily be removed from the circuit and a new one substituted. The Shreeve repeater was, relatively speaking, widely used until the advent of the vacuum-tube repeater about 1913.

Vacuum-Tube Telephone Repeaters.[12] As indicated in Fig. 4, a vacuum-tube telephone repeater consists of two parts: the vacuum-tube amplifier (page 285) and the associated circuit and equipment. The repeaters used for voice-frequency circuits employ either triodes or pentodes in class A (page 291). In the carrier-frequency circuits the repeaters are often pentodes in class A feedback amplifier circuits (page 300). Only voice-frequency repeaters will be discussed in the pages immediately following.

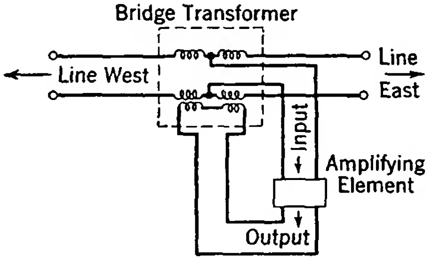


FIG. 4. A simple two-way repeater using a hybrid coil.

The Hybrid Coil.[1] This hybrid coil is a special transformer having three sets of windings, although only two sets will be considered in the following discussion. The hybrid coil, also called a **bridge transformer,** operates on the principle of the impedance bridge. In Fig. 5(*a*), if corresponding elements are identical the bridge will be balanced, the receiver terminals will be at the same potential, and no sound will be produced by the receiver. If now the circuit is rearranged as in

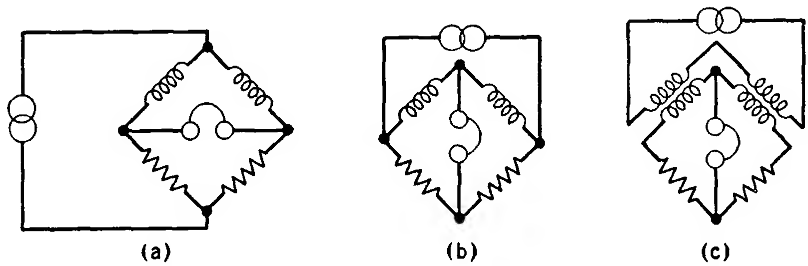


FIG. 5. Circuits for developing the principle of operation of the hybrid coil from the theory of the impedance bridge.

Fig. 5(*b*), no difference of potential will exist across the receiver terminals, and the bridge will still be balanced. Introducing the energy into the bridge circuit as in Fig. 5(*c*) does not alter the condition of balance.

The last bridge circuit has been redrawn in Fig. 6. In this figure, Z_w and Z_e represent the impedances of the lines west and east, respectively, and Z_i and Z_o represent the impedances of the input and output circuits of the amplifying element of Fig. 4.

When an impulse comes in over line west of Fig. 6, it is impressed

across the input Z_i of the element and the amplified signal is introduced on the line by the output coil Z_o of the transformer (see also Fig. 4). *This amplified impulse must not again introduce energy into the input circuit or continuous oscillations will be produced and the repeater will sing or howl.* The action and requirements are similar for an impulse from line east. There are then three sources of voltage to consider: *first,* a voltage impressed at Z_w as in Fig. 6(*a*); *second,* a voltage (not shown) impressed at Z_e; and *third,* a voltage impressed at Z_o as in Fig. 6(*b*).

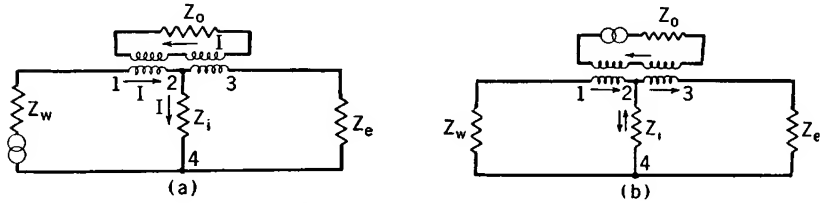


FIG. 6. The hybrid coil (or bridge transformer, as it is sometimes called) is connected as shown, except that in practice a balanced hybrid coil with additional line windings would be used (Figs. 4, 7, and 8).

In Fig. 6(*a*) the voltage impressed in series with Z_w represents an incoming impulse from line west. Assume that Z_e is temporarily disconnected, leaving terminals 3–4 open; then the impressed voltage will cause a current flow as indicated by the arrows. If the value of Z_o and the turns ratio are such that the impedance measured *between terminals* 1–2 (with the remainder of the circuit disconnected) equals Z_i, then the opposing voltage drop across the coil 1–2 will equal that across Z_i. The identical coil 2–3 is linked by the same flux as coil 1–2 and will therefore have a voltage induced in it equal to and in the same direction as the voltage across coil 1–2. This *induced* voltage will also be equal to the voltage drop Z_i. Thus, points 3 and 4 are at the same potential, and no current will flow through Z_e when line east is again connected. For the ideal conditions assumed, one-half of the incoming energy from line west is dissipated in Z_o, and the remainder is available in the input circuit Z_i for actuating the amplifying element. The action is similar for a signal voltage at Z_e instead of Z_w.

Fig. 6(*b*) represents the output of the amplifying element being induced into the telephone line. Coils 1–2 and 2–3 are connected in series *aiding,* and one-half of the total voltage will be induced across each. Currents will flow at a given instant as indicated by the arrows. The voltage drops across the equal impedances Z_e and Z_w will be the same, and thus no difference of potential will exist across Z_i and no

current will flow through the amplifier input as indicated by the opposing current arrows at Z_i. Thus, the output circuit will *not* feed back into the input circuit and will not cause sustained oscillations.

The bridge transformer has other uses in communication and is also adapted to numerous measuring circuits. Thus, with a voltage impressed at Z_o, differences in impedances connected at Z_e and Z_w are readily detected by a tone heard in the headphones connected at Z_i. A mathematical treatment of the hybrid coil is given in reference 14.

The 21-Type Repeater.[12, 15] The circuit of this repeater, shown in Fig. 7, is fundamentally the arrangement of Fig. 6. To explain the operation of this circuit, assume that a voice-frequency signal comes in from line west. If the impedances of the input and the output circuits are equal, the incoming energy will divide, half being dissipated in the output circuit and the other half being introduced into the input circuit where it is amplified. This amplified power is again introduced on the line, where it divides, one half flowing west where the speaker hears his voice as an echo, and the other half flowing to the listener on the east line.

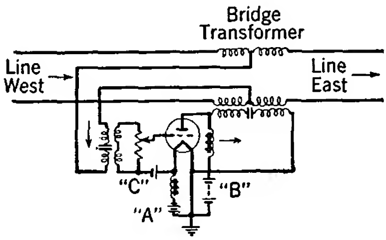


FIG. 7. Simplified diagram of a 21-type repeater.

The 21-type repeater is very simple in operation, requires but a small amount of equipment, is relatively cheap to install and operate, but *is limited in application.* To prevent oscillations, or **singing,**[1] the impedance of line west must equal that of line east, for, as the preceding section shows, if these relations do not hold the output and the input circuits will be coupled.

Because the 21-type repeater sends amplified energy in both directions, the repeater is not suited for tandem operation (several repeaters installed at different points in the same line) because the energy flowing back toward the speaker (and toward the listener) would be reamplified and reintroduced in both directions on the line at each repeater point, causing bothersome echoes.

The 22-Type Repeater.[12, 15] These limitations are largely overcome in the 22-type repeater shown in Fig. 8. Two amplifying elements are used, and each line is balanced by a network. Thus, lines of different characteristics may be connected to the repeater without causing singing.

Assume that a voice-frequency signal comes in over line west. The

energy will divide at the bridge transformer, one half being dissipated in the output circuit of element 2 and the other half entering the input circuit of element 1. This energy is then amplified and introduced into line east, where it divides, one half being lost in the balancing network and the other half flowing over the line to the distant listener. Thus, in order for a 22-type repeater to sing, an unbalance must exist between each line and its balancing network; that is, the west bridge transformer must be unbalanced to introduce energy into the input of

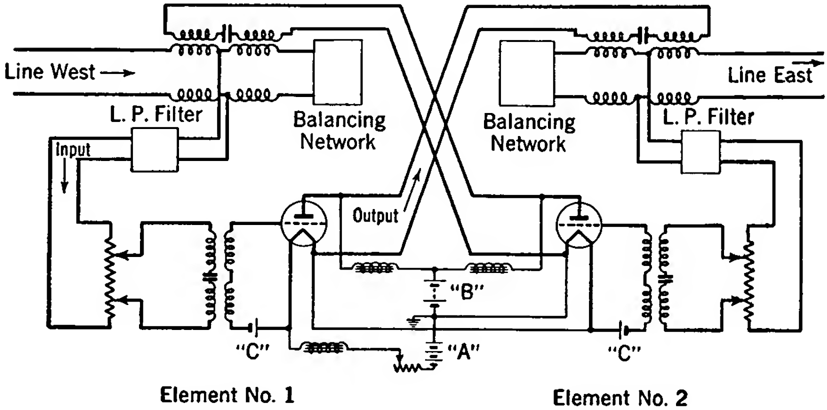


FIG. 8. Simplified diagram of a 22-type repeater.

element 1, and the east transformer must be unbalanced to permit energy flow into the input of element 2 and again couple with the west transformer. The purpose of the low-pass filter shown in Fig. 8 is to eliminate currents above the voice-frequency range, thus making it unnecessary for the balance between the line and the network to hold accurately at high frequencies. This amplifier has been widely used in telephone systems.

Four-Wire Repeater.[2, 16] Often it is desirable to use **four-wire circuits**[1] for telephone toll purposes. With this method, talking is in one direction over one pair of wires, and in the other direction over another pair. Such an arrangement is especially suited to toll cables; simpler repeaters can be used, and these can be operated at higher gains than with two-wire circuits, thereby necessitating fewer repeaters. Although twice as many cable wires are used, they may be smaller, and this fact, combined with the advantages of fewer and simple repeaters, often makes the four-wire circuit economical to use.

A block diagram of the four-wire arrangement including four-wire **44-type repeaters** is shown in Fig. 9. It will be noted that at each

repeater point only *one-way repeaters* are required and that hybrid coils and balancing networks are not necessary. Hybrid coils and balancing networks reduce the four-wire circuit to a two-wire circuit at each end. Four-wire repeaters are worked at high gains, and thus the conductors transmitting in opposite directions must be separated as much as possible to prevent crosstalk between the high- and low-energy level cable pairs. Special cable arrangements and office wiring are used, and the circuits are *segregated* into the groups and separated.

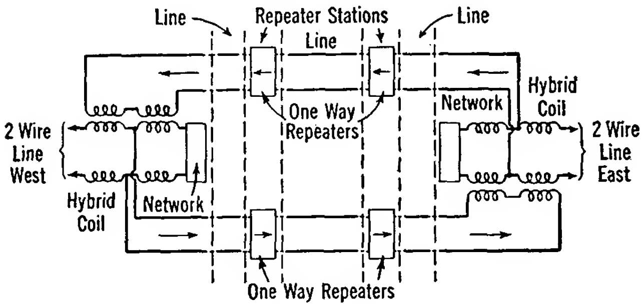


FIG. 9. A four-wire circuit, equipped with 44-type telephone repeaters and arranged for connection to two-wire circuits at the terminals.

The V1 Repeater.[17] The basic 22-type repeater and the modifications 22A1 and 22A2 were standard from about 1915 to 1940. In these, hybrid coils, balancing networks, filters, etc., which in a sense are as much line equipment as repeater equipment, were associated with the amplifying elements. About 1940 the **V1 repeater** was perfected. This is a voice-frequency repeater in which the line equipment *is physically separated* from the amplifying units, giving increased flexibility, reduction in space requirements and cost, and other advantages.

A simplified diagram of the V1-repeater arrangement is shown in Fig. 10. The center portion, consisting of the amplifiers, is mounted, together with amplifiers of other similar repeaters, in one location. The two end portions, consisting of line equipment, are at another location. Of importance is the fact that two transformers, or repeating coils, instead of a hybrid coil, are used to connect to the incoming circuits. The operation is summarized[17] as follows:

The + and − signs adjacent to the windings of the repeating coils indicate the relative "poling" of the windings; that is, they indicate the relative directions of the induced voltages. A speech signal coming from line west passes through coils A1 and B1 and produces equal magnetic fluxes in the cores of each transformer. The flux in the core of

transformer A induces voltages in windings A2 and A3, and the flux in the core of transformer B induces voltages in windings B2 and B3. The voltages induced in A2 and B2 are opposite in sign, and the design is such that there is complete cancellation and no current flow in the balancing network if the impedances connected to A3 and B3 are equal.

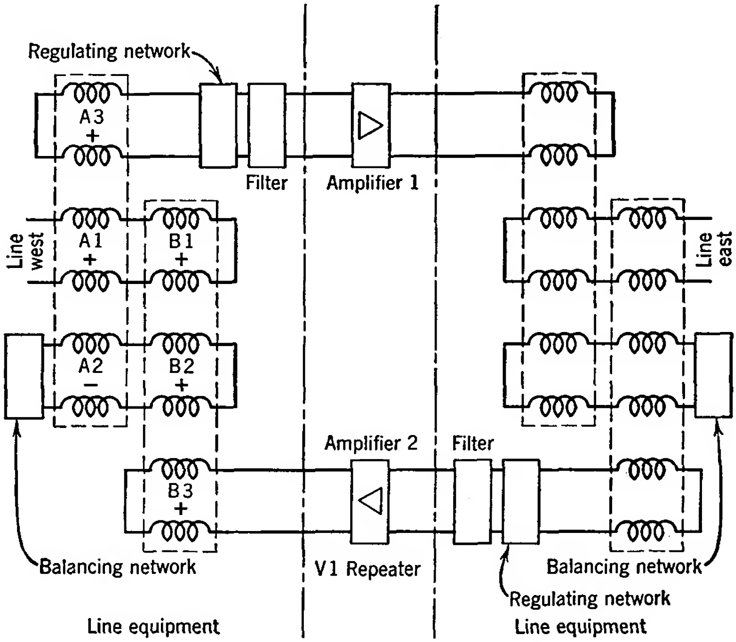


FIG. 10. Simplified diagram of the V1-repeater arrangement. Note that the transformers, networks, and filters are associated with the line and that the V1-repeater consists of the amplifiers. The four transformers are shown enclosed with broken lines. (Reference 17.)

The energy coming in over line west divides equally between the input to the upper branch, where it is amplified and passed to the line east at the right, and the output to the lower branch where it is dissipated.

An amplified speech signal, such as that coming from the lower amplifier passes through winding B3 and produces a magnetic field in the core of transformer B. This magnetic field induces voltages in the B1 and B2 windings, and these voltages cause current to flow to line west and to the associated balancing network. If the line impedance and the network impedance are equal, these currents produce no magnetic flux in the core of transformer A, because windings A1 and A2 are opposite in their magnetic action. Thus, no current flows through the

A3 winding and into the input of the upper amplifier, preventing singing. The signal energy from the lower amplifier divides equally between line west and the associated balancing network, and none goes to the upper amplifier. Later developments in repeaters have retained the basic V1-repeater principles.

It appears possible (1949) that repeaters of the future may be affected by the development of the Transistor (page 311).

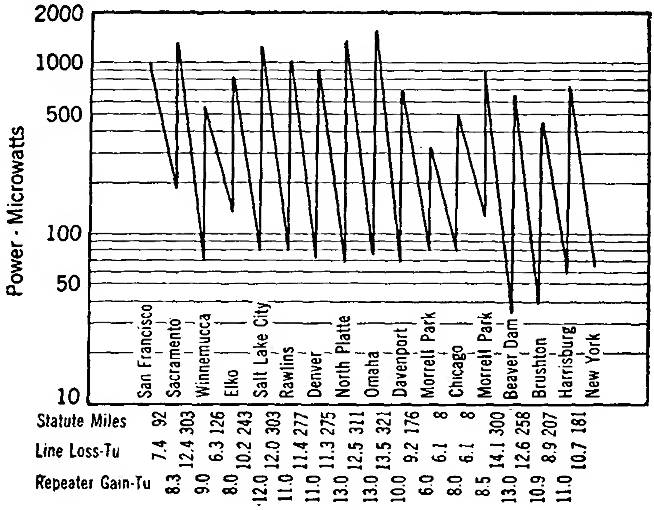


FIG. 11. Power levels of San Francisco-New York *open-wire* connection at 1000 cycles, when power of 1000 microwatts is applied at San Francisco. (Although this particular voice-frequency circuit is not used at present, this illustration is retained because of its simplicity and historical interest. The loss and gain in Tu refers to an old unit of measure, the transmission unit, which has been superseded by the decibel.) (Reference 18.)

Repeater Applications. The function of a telephone repeater is to amplify the weakened currents at various points along the line or cable. The manner in which this *was* accomplished for an early transcontinental *open-wire* connection is illustrated by Fig. 11. It may appear that all the amplification could be supplied at the sending end instead of at various points along the line. That this is impracticable is shown by the following statement:[18]

Let us suppose, however, that a 50,000-kilowatt generator delivered its entire output to the circuit at San Francisco, and overlook, for the moment, what would happen to the line if any such amount of energy were applied. The power received at New York would be of the order of one five-hundredth

of a microwatt, which would have to flow for about 25,000 years in order to equal the energy required to light a 25-watt lamp for one minute.*

It may seem possible to do all the amplifying at the receiving end. This cannot be done because the ratio of the useful speech currents to the undesired noise current level must always be kept high. If at any time the level of the speech currents fall below the noise currents, then at the receiver the noise will be greater than the speech. Repeaters *must* be installed at regular intervals along a line. In Fig. 12 is shown the loss at various frequencies over a 30-mile loaded cable section. This

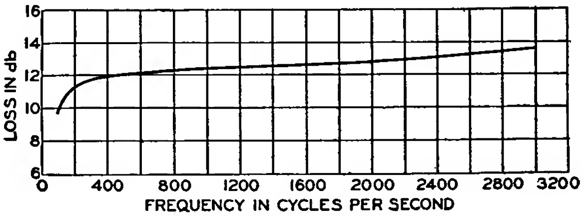


Fig. 12. Loss-frequency curves of a 30-mile section of a 16-gauge, H–44–25 loaded cable circuit. (Reference 15.)

loss, it is seen, varies with frequency. Not only does the repeater amplify the speech current and overcome the loss, but it also corrects the frequency distortion.[15]

On long aerial toll cables special provisions must be made to compensate for changes in attenuation caused by temperature variations affecting the resistance and to some extent the capacitance of the circuits.[19] This is sometimes accomplished by a Wheatstone-bridge arrangement. Two vacant wires in the cable are connected to form one side of a bridge. The change in resistance of these wires operates the bridge mechanism, which in turn changes the repeater amplification to compensate for the greater or less line attenuation.

Repeaters are applied to voice-frequency circuits in several forms. First, they are installed at various points along an open-wire line or cable circuit, in which event they are called[1] **intermediate repeaters,** or sometimes **through-line repeaters.** Second, they are installed at the ends of transmission circuits and are hence called[1] **terminal repeaters.** Third, repeaters may be installed in the cord circuit of a toll switchboard and are hence called **cord-circuit repeaters.** The toll operator uses this particular cord circuit for interconnecting toll lines having high loss. Formerly this practice was widely used, but in

* Reprinted by permission, courtesy H. H. Nance and O. B. Jacobs, and the American Institute of Electrical Engineers.

modern systems most toll circuits are provided with repeaters (if necessary) and are so designed that for all possible connections the loss is not excessive.[3]

Echo Suppressors. It is not practicable to remove *all* impedance irregularities from telephone circuits, and those remaining may under certain conditions cause wave reflections producing serious "echo" effects.[20, 21] To be troublesome, the reflected wave must return after a certain perceptible time interval, and hence two factors are involved. Troublesome echoes are most likely to occur in *long* circuits where the *time* required to travel is appreciable and in loaded cables where the velocity of propagation is *low* (page 246). In fact, the low velocity obtained with heavy loading limits its use.

An early **echo suppressor,**[1] consisted of two high-impedance vacuum-tube amplifier-detectors bridged across the line. Each amplifier-detector was connected to a relay, and, when an alternating speech voltage of sufficient strength was impressed across the detector, the relay was operated, placing a short circuit on the opposite line. This short circuit prevented reflected echo waves from traveling back on the other side of the four-wire circuit to the speaker.[22]

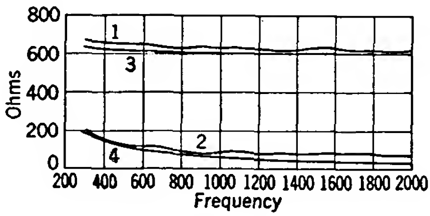


FIG. 13. Curve 1 is the resistance component, and curve 2 is the capacitive reactance component of the input impedance of a typical open-wire line terminated in its characteristic impedance. Curve 3 is the resistance component, and Curve 4 is the capacitive reactance component of the input impedance of the balancing network of Fig. 14(*b*). The frequency is in cycles per second.

Balancing Networks. As explained on page 404, the impedance offered by a line or cable connected to a hybrid coil in a telephone repeater must be balanced by the impedance of a network also connected to the hybrid coil. For the usual open-wire telephone line, the impedance offered to the hybrid coil will be approximately the characteristic impedance of the line. For a loaded cable, the impedance offered to the hybrid coil will depend on the type of cable, type of loading, and the terminating sections used (page 248).

Impedance variations for an open-wire telephone line composed of 165-mil hard-drawn copper wires spaced one foot apart are shown in Fig. 13. Curve 1 is the resistive component of the input impedance, and curve 2 is the capacitive reactance component of the input impedance, with the distant end properly terminated.[12] Theoretically,

these curves should be perfectly smooth, but small irregularities cause minor reflections and introduce the small impedance variations.

It is evident that the resistance component is by far the larger of the two, and hence the line could be approximately balanced with a 600-ohm resistor. Accurate balancing, such as in repeater circuits, demands a closer simulation, and this can be accomplished by adding a capacitor in series with the resistor as is done in Fig. 14(*a*). The network can be made to balance the line more exactly by adding another capacitor and another resistor as in Fig. 14(*b*). The curves for this final balancing network are numbers 3 and 4 of Fig. 13, and as is evident they closely balance the line impedance over a wide range of frequencies.

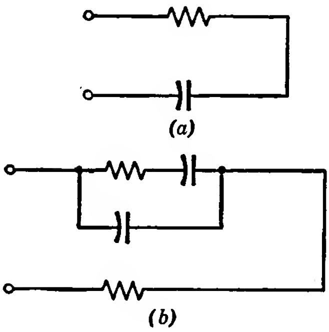


FIG. 14. Simple balancing networks sometimes used with open-wire lines.

The balancing networks just described simulate the impedance of an open-wire line. Balancing networks[23, 24] are also needed for use in repeaters connected to loaded cables.

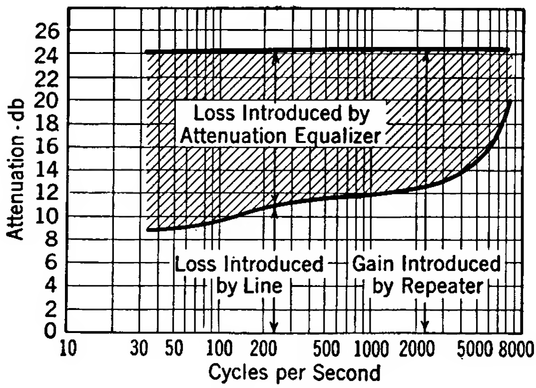


FIG. 15. Attenuation-frequency characteristics of line equalizer and 50 miles of 16-gauge, B-22 cable circuit. (Reference 25.)

Attenuation Equalizers.[23, 24] Communication circuits usually offer greater attenuation at certain frequencies than at others. This is well illustrated by Fig. 15, which shows the losses offered by a cable circuit.[25] Since this circuit was for radio program networks, the distortion would be excessive.

To compensate for this distortion, an **attenuation equalizer**[1] was inserted in the line. This equalizer had characteristics which are complementary to those of the line, and thus, when the loss in the line and that in the attenuation equalizer are combined, the overall loss is almost the same at all frequencies, as is illustrated by Fig. 15. The total attenuation is greater than that due to the line alone, but this can easily be compensated for by repeaters.

Delay Equalizers.[26] In high-quality circuits, such as radio program networks, it is desirable that all frequencies travel along a cir-

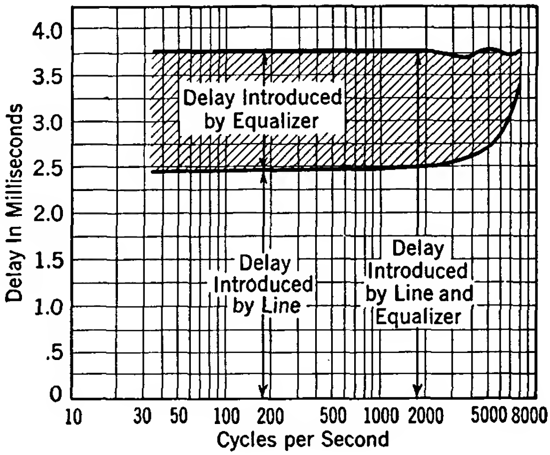


FIG. 16. Delay-frequency characteristics of 50 miles of 16-gauge, B-22 cable with and without delay equalizer. (Reference 25.)

cuit at the same velocity or the program will be distorted. Fig. 16 illustrates this effect in a loaded radio program cable. It will be observed that for the cable alone the lower frequencies travel faster than the high-frequency components. When a **delay equalizer**[1] is inserted in the line, however, the overall velocity of transmission is the same, and distortion from this cause is prevented.

Carrier Telephony.[27] Early in the history of telephony efforts were made to develop methods for transmitting more than one telephone conversation over a pair of wires. This led to the use of **carrier transmission,** defined[1] as "that form of electric transmission in which the transmitted electric wave is a wave resulting from the modulation of a single-frequency wave by a modulating wave." Carrier transmission is employed in systems of **carrier telephony,** defined[1] as "that form of telephony in which carrier transmission is used, the

modulating wave being a voice-frequency wave." Ordinarily, the term is applied only to wire telephony. Carrier telephone systems were used first in about 1918.

The principle of operation of the usual multichannel carrier telephone system is as follows: For *each* channel the incoming voice-frequency signals are used to amplitude-modulate a carrier wave of higher frequency. As will be explained (page 413), the modulation results in the creation of two sidebands, each of which is at a higher frequency than the voice-frequency signals, and *each* of which contains all the signal variations necessary for the transmission of speech.

In most carrier telephone systems *one sideband only* for each channel is transmitted *simultaneously* with sidebands from other channels over the same pair of wires or cable circuit. Carrier repeaters are used to offset circuit attenuation.

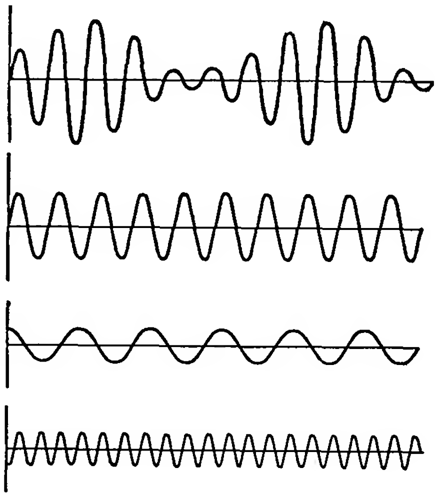


FIG. 17. An amplitude-modulated carrier wave *a* is composed of a carrier-frequency component *b*, a lower sideband *c*, and an upper sideband *d*. The modulating wave used was a single frequency of constant amplitude. (These curves are only approximate.)

At the receiving end the signals are separated by filters, and each signal is demodulated by a process fundamentally the same as that of modulation. As a result, the desired voice-frequency component is created from the otherwise unintelligible carrier signal. The voice-frequency component is transmitted over the subscriber line to the listener.

Analysis of an Amplitude-Modulated Wave.[28] By definition[1], **modulation** is the "process whereby the amplitude (or other characteristic) of a wave is varied as a function of the instantaneous value of another wave. The first wave, which is usually a single-frequency wave, is called the **carrier wave;** the second wave is called the **modulating wave.**" In **amplitude modulation** the amplitude of the impressed carrier wave is caused to vary in accordance with the voice-frequency modulating wave to be transmitted.

A portion of an amplitude-modulated output wave (of the type

transmitted in radio-broadcast systems, page 489) is shown in Fig. 17(*a*). The modulating signal is a sinusoidal wave, such as a 1000-cycle tone.

The instantaneous magnitude of a sinusoidal current wave is

$$i = I_m \sin \omega t. \tag{1}$$

In amplitude modulation, the value of I_m of a carrier-frequency wave $\omega_c = 2\pi f_c$ is caused to vary in amplitude by a voice-frequency signal wave of frequency f_v in accordance with the relation

$$I_m = I_c + mI_c \sin \omega_v t, \tag{2}$$

where I_c is the maximum value of the unmodulated carrier wave, **m** is the **percentage modulation,** and $\omega_v = 2\pi f_v$.

Substituting equation 2 in equation 1 and writing ω_c for ω gives, as the instantaneous value of an amplitude-modulated wave such as Fig. 17(*a*),

$$i = I_c \sin \omega_c t + mI_c \sin \omega_v t \sin \omega_c t. \tag{3}$$

From trigonometry, $\sin \omega_v t \sin \omega_c t = \frac{1}{2} \cos (\omega_c - \omega_v)\, t - \frac{1}{2} \cos (\omega_c + \omega_v) t$, and if these substitutions are made equation 3 becomes

$$i = I_c \sin \omega_c t + \frac{mI_c}{2} \cos(\omega_c - \omega_v)t - \frac{mI_c}{2} \cos(\omega_c + \omega_v)t. \tag{4}$$

This is the equation for an amplitude-modulated wave such as shown in Fig. 17(*a*), and the equation shows that this wave contains *three* components as follows: *first,* a carrier-frequency component, $I_c \sin \omega_c t$, which is not affected by the modulation process and is shown by Fig. 17(*b*); *second,* the **lower sideband,**[1] $\frac{mI_c}{2} \cos(\omega_c - \omega_v)t$, of maximum amplitude $mI_c/2$ and frequency $\left(\frac{\omega_c}{2\pi} - \frac{\omega_v}{2\pi}\right)$ or $(f_c - f_v)$, and as shown by Fig. 17(*c*); *third,* the **upper sideband,**[1] $\frac{mI_c}{2} \cos(\omega_c + \omega_v)$, of maximum amplitude $mI_c/2$ and frequency $\left(\frac{\omega_c}{2\pi} + \frac{\omega_v}{2\pi}\right)$ or $(f_c + f_v)$, and as shown by Fig. 17(*d*). *It is important to note that, in amplitude modulation, sum frequencies* $(f_c + f_v)$ *and difference frequencies* $(f_c - f_v)$ *are created.*

The preceding analysis and Fig. 17 are for amplitude modulation by a *single-frequency* wave. In this instance the sidebands (more correctly, **side frequencies**) also will be single-frequency waves. For example, if the carrier frequency is 10,000 cycles, and the modulating

frequency is 1000 cycles, the lower side frequency will be a single frequency of 9000 cycles, and the upper side frequency will be a single frequency of 11,000 cycles.

If the carrier frequency is 10,000 cycles and the modulating frequency is a telephone conversation that is contained within a band of from 200 to 3500 cycles, then the lower sideband will be a complex wave lying within the band 9800 to 6500 cycles, and the upper sideband will lie within the band 10,200 to 13,500 cycles. It is important to note that in amplitude modulation *two* sidebands are created and that *each* of these varies in accordance with the modulating signal and contains *all* the information of the telephone conversation to be transmitted. The component *of carrier frequency* is unchanged by the amplitude-modulation process and contains no information. The idea that the component of carrier frequency, or "carrier," as it is usually called, in some way "carries" the information to the distant station is incorrect. In fact, most carrier telephone systems and many radio systems suppress the carrier and one sideband and transmit only one sideband to the distant station.

Types of Carrier Telephone Systems. Some of the systems that have been devised are listed in Fig. 18. This chart does *not* include all models of certain systems. The type A and type B systems are no longer used but are of fundamental interest.

The various carrier systems have features in common, yet each has distinguishing characteristics. They provide excellent telephone channels, and their importance in comparison with other facilities is indicated by Fig. 19. The types C, J, and K systems listed in Fig. 18 and the type L system are of most importance and will be considered in greatest detail.

Type A Carrier Telephone System.[27] The original carrier used by the Bell System was designated type A. The carrier component and one sideband were suppressed, and one sideband from each channel was transmitted. In this system *four* talking channels, operating at different frequencies, were provided over one open-wire pair. The *same* frequencies were used for talking in each direction as indicated by the double-headed arrows of Fig. 18. Because of this it was necessary to use a carrier-frequency hybrid coil to separate the "talking" and "receiving" sidebands. This hybrid coil and the required balancing networks, which operated at carrier frequencies, were objectionable features. The type A system transmitted a master 5000-cycle frequency throughout the system. This was distorted at the various carrier terminals, and the harmonics were selected by filters and amplified to supply the carrier frequencies necessary for modulation (page 418)

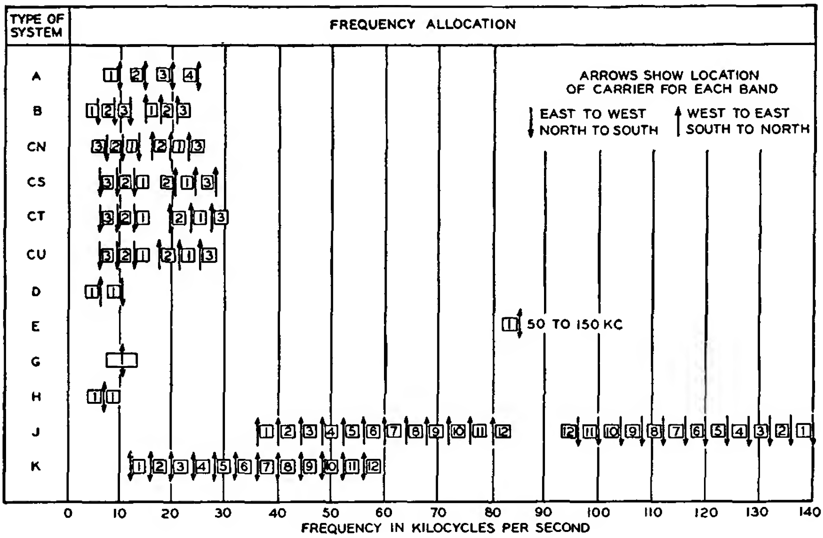


FIG. 18. Approximate frequency allocations of certain carrier telephone systems used in the United States. The improved type C5 carrier uses the CS and CN allocations. The type E is a power-line carrier system. Two systems not shown are the type L (page 432), used on coaxial cables, and the type M (page 434), used for giving service over rural power lines and for other purposes (page 434). Of the systems shown, only the types B and G transmit the carrier component.

and demodulation (page 421). Several type A systems were installed about 1918, one of which remained in service until about 1940.[29]

Type B Carrier Telephone System.[27] The second carrier, developed and installed about 1920, was designed type B; one set remained in operation until about 1940.[29] The carrier component and one sideband were transmitted. Because of this and the fact that stable oscillators had been developed, it was no longer necessary to

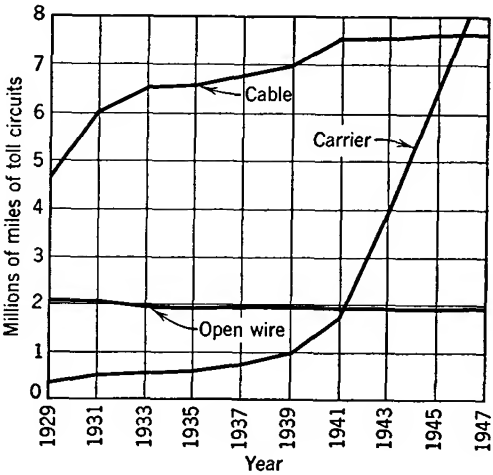


FIG. 19. Approximate circuit miles of voice-frequency telephone facilities on open-wire toll lines and cables as compared with circuit miles of carrier-frequency toll facilities.

transmit a master 5000-cycle control frequency throughout the system.[27] Also, as Fig. 18 indicates, different frequencies were used for "talking" and "receiving," or for "east-west" and "west-east" transmission. Thus, the "talking" and "receiving" sidebands could be separated by filters, and carrier-frequency hybrid coils and balancing networks were not necessary. With the type B system, both the transmitted carrier and the sideband were attenuated by the line. Because the magnitude of the demodulated voice-frequency wave depends on the magnitude of *both* the carrier and the sideband, the voice-frequency wave was decreased by the *square* of the attenuation. Variations in transmission, caused by weather changes, were pronounced.[27]

Type C Carrier Telephone Systems. The type C system was first installed about 1925 and has been very extensively used. It incorporates the most desirable features of the systems just considered. Like the type A system, the carrier is suppressed, and hence the effect of line attenuation is minimized. Like the type B, different frequencies are used for transmitting and receiving, and filters, instead of carrier-frequency hybrid coils, are used to separate the talking and the receiving signals. The original type C carrier telephone system was extensively modified about 1938 to take advantage of developments in communication.

Original Type C Carrier Telephone System.[30] The schematic diagram of the original type C carrier is shown in Fig. 20. To explain the operation of this system, assume that a person whose telephone is connected to channel 1 at the left is speaking to a person whose telephone is connected to channel 1 at the right.

The incoming speech signals pass through the *voice-frequency* hybrid coil, and to the modulator at the top. This is a balanced modulator (page 421) that eliminates the carrier but produces both sidebands. The band filter selects one sideband and rejects the other. The selected sideband passes down to the common wires leading to the transmitting amplifier, where all the incoming signals to be transmitted are amplified to the proper level for impressing on the line. The amplified signals from the three channels pass through the high-pass filter and hence to the line. Each of the three modulators is fed with a different carrier frequency, and hence each of the three sidebands that are transmitted simultaneously is at a different frequency as Fig. 18 shows.

At a repeater point the carrier frequencies are separated from the voice frequencies by the high-pass and low-pass filters. At the carrier amplifier the transmitting and the receiving signals are separated by

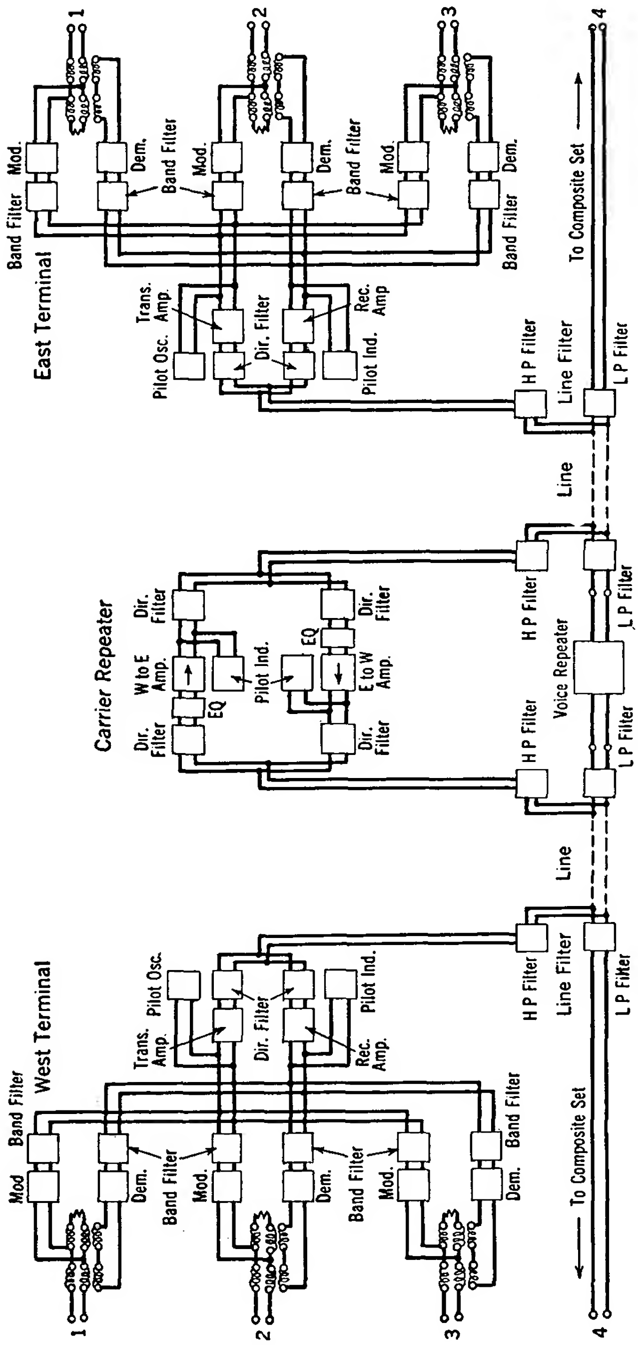


FIG. 20. Schematic diagram of original type C carrier telephone system, and of carrier and voice-frequency equipment at a repeater station. (Reference 30.)

directional filters. The equalizer functions as explained on page 410 The sideband signals of the three channels are amplified in the "west to east" amplifier,[30] pass through the directional filter, the high-pass filter, and the amplified signals are then impressed on the line toward the "east terminal" at the right of Fig. 20.

At the east terminal the received sidebands from the transmitting "west terminal" are passed by the high-pass line filter upward to the directional filters. Since the east terminal now is being considered as receiving from the west terminal, the three incoming sidebands will pass through the lower directional filter and to the common circuit leading to the demodulating apparatus. The sideband from channel 1 will be passed by the band-pass filter of channel 1 to the demodulator (page 421). The output of the demodulator will contain the original message at audio frequencies, and this is passed by the voice-frequency hybrid coil to the telephone set of the listener. Channels 2 and 3 function in a manner similar to that just explained. The pilot channel noted on Fig. 20 is for monitoring and adjusting the transmission to ensure that the signal strength is correct at all times.[30]

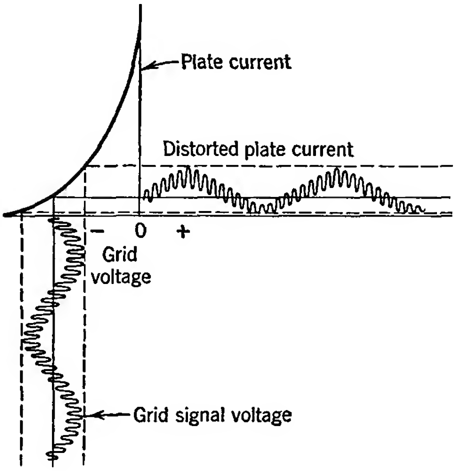


Fig. 21. This illustrates the principle of modulation (and demodulation) that was used in early telephone carrier systems employing vacuum tubes in the modulators (and demodulators). This is known as the van der Bijl method. Distortion and the creation of sidebands occur in the plate circuit.

Modulation in the Original Type C Systems.[30] As discussed in the preceding paragraphs, sidebands, which contain the information to be transmitted, are created in the modulator. A **modulator** is defined[1] as "a device to effect the process of modulation. It may be operated by virtue of some non-linear characteristic, or by a controlled variation of some circuit quantity." Modulation was defined on page 412.

In the original type C carrier system the non-linear relation between grid voltage and plate current in triode thermionic vacuum tubes was used to produce modulation. To accomplish this, the speech signals to

be transmitted and the carrier frequency are impressed simultaneously on the grids of tubes so biased that operation is on the non-linear portion of the characteristic curves as shown in Fig. 21. As will be explained, the sidebands are *created in the process of distortion.* This statement has been italicized because sometimes it is thought that

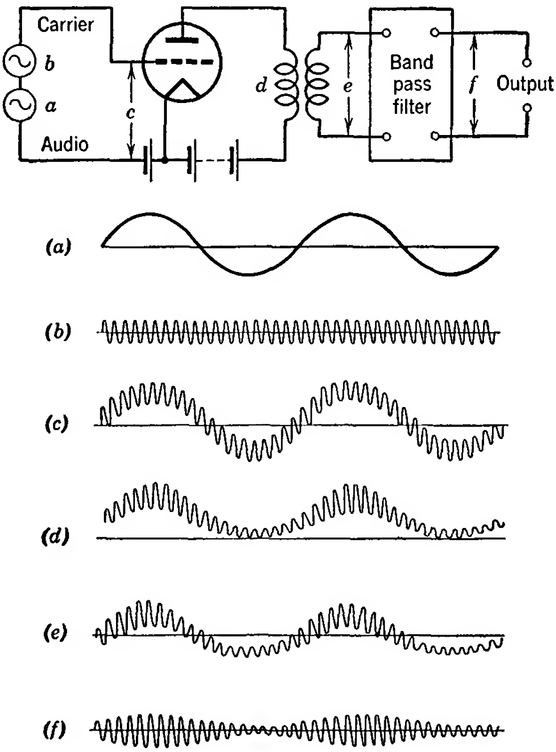


FIG. 22. Circuit for the van der Bijl system of modulation of Fig. 21, and analysis of currents and voltages in various parts of the circuit. (See also Fig. 17.)

merely "mixing" two waves, or "beating" two waves, creates sidebands. This is *not true:* the sidebands are created by simultaneously *distorting* the two waves, and the modulating tubes must be operated so that they distort, and not merely amplify, if the sidebands are to be created.

If a *single* tube is operated so that it distorts as shown in Fig. 21, the shapes of the resulting waves will be as shown in Fig. 22. As indicated, the final signal is a modulated wave containing the carrier and two sidebands as explained on page 413.

Over the curved portion used in modulation as shown in Fig. 21, the

equation for the plate current is approximately

$$i_b = ae_g + be_g^2 + \cdots. \tag{5}$$

In this equation a and b are constants, and e_g is the instantaneous grid voltage. When the instantaneous input carrier voltage wave is $E_c \sin \omega_c t$ and when the instantaneous modulating voltage wave is $E_v \sin \omega_v t$, then

$$e_g = E_c \sin \omega_c t + E_v \sin \omega_v t. \tag{6}$$

When this expression is substituted in equation 5, the instantaneous plate current becomes

$$i_b = aE_c \sin \omega_c t + aE_v \sin \omega_v t + bE_c^2 \sin^2 \omega_c t + bE_v^2 \sin^2 \omega_v t + 2bE_cE_v \sin \omega_c t \sin \omega_v t. \tag{7}$$

From trigonometry,

$$\sin \omega_c t \sin \omega_v t = [\tfrac{1}{2} \cos (\omega_c - \omega_v)t - \tfrac{1}{2} \cos (\omega_c + \omega_v)t].$$

Also,

$$\sin^2 \omega_c t = (\tfrac{1}{2} - \tfrac{1}{2} \cos 2\omega_c t), \quad \text{and} \quad \sin^2 \omega_v t = (\tfrac{1}{2} - \tfrac{1}{2} \cos 2\omega_v t).$$

If these substitutions are made, equation 7 becomes

$$\begin{aligned} i_b = aE_c \sin \omega_c t + aE_v \sin \omega_v t &+ bE_c^2(\tfrac{1}{2} - \tfrac{1}{2} \cos 2\omega_c t) \\ &+ bE_v^2(\tfrac{1}{2} - \tfrac{1}{2} \cos 2\omega_v t) \\ &+ 2bE_cE_v[\tfrac{1}{2} \cos (\omega_c - \omega_v)t - \tfrac{1}{2} \cos (\omega_c + \omega_v)t]. \end{aligned} \tag{8}$$

When the indicated multiplications are performed and all constant terms dropped (since the alternating portions only are of primary interest), equation 8 may be written

$$\begin{aligned} i_b = aE_c \sin \omega_c t + aE_v \sin \omega_v t &- \frac{bE_c^2}{2} \cos 2\omega_c t - \frac{bE_v^2}{2} \cos 2\omega_v t \\ &+ bE_cE_v \cos (\omega_c - \omega_v)t - bE_cE_v \cos (\omega_c + \omega_v)t. \end{aligned} \tag{9}$$

Thus, remembering that $f_c = \dfrac{\omega_c}{2\pi}$ represents the carrier frequency, and that $f_v = \dfrac{\omega_v}{2\pi}$ represents the voice frequency, the following alternating components exist in the modulator output:

1. f_v (the voice frequency, represented by the second term).
2. $2f_c$ (twice the carrier frequency, represented by the third term).
3. $2f_v$ (twice the voice frequency, represented by the fourth term).
4. f_c (the carrier frequency, represented by the first term).
5. $f_c - f_v$ (the lower *sideband*, represented by the fifth term).
6. $f_c + f_v$ (the upper *sideband*, represented by the last term).

If the first three components are eliminated (with the filters), then the resulting wave will have the shape of Fig. 22(*f*) and will have the characteristics discussed on page 414. In the type C carrier system the carrier and one sideband are suppressed. The carrier is suppressed in the push-pull **balanced modulator** of Fig. 23. In this circuit the carrier voltage drives the two grids in phase, but the voice-frequency signal voltage drives the two grids 180° out of phase. The unwanted sideband is suppressed by the band-pass filter associated with each modulator in Fig. 20.

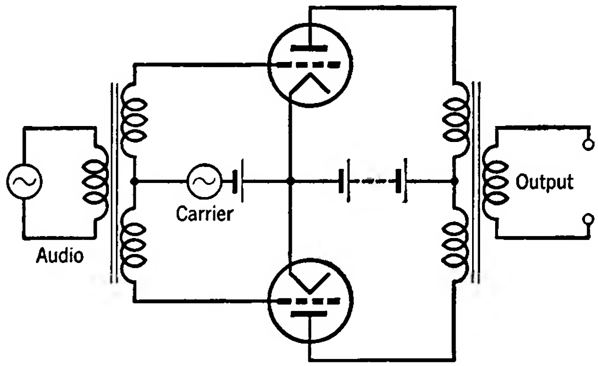


FIG. 23. A balanced modulator arranged for suppressing the carrier in a carrier-telephone system.

The statements made regarding cancellation of the carrier can be proved as follows: Let the instantaneous carrier voltage be expressed as before, $E_c \sin \omega_c t$. Let the instantaneous voice-frequency modulating voltage be $+E_v \sin \omega_v t$ for one tube, and $-E_v \sin \omega_v t$ for the other tube. Expansions are then made for the plate current of each tube as previously explained for one tube, and the expressions are *subtracted,* because each current flows through one half of the primary in a direction opposite to that of the other.

Demodulation in Original Type C Carrier Systems.[30] **Demodulation** is defined[1] as "the process whereby a wave resulting from modulation is so operated upon that a wave is obtained having substantially the characteristics of the original modulating wave."

The demodulating circuit of the original type C carrier system is essentially the circuit of Fig. 23 used for modulation. Because the carrier frequency is not transmitted, it must be supplied by an oscillator associated with the demodulator. Studies similar to those made for modulation will show that, if the received sideband and the locally

generated carrier are impressed on the circuit (Fig. 22), a difference frequency will be created. This is the desired voice-frequency signal wave, having the characteristics of the original modulating wave and containing the information that has been transmitted.

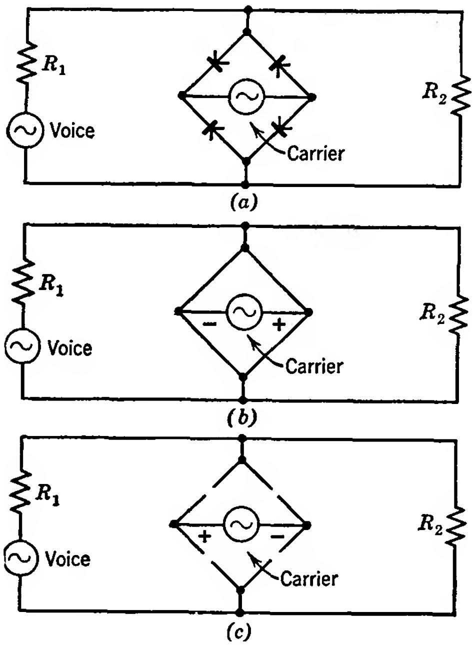


FIG. 24. This figure illustrates the principle of modulation and demodulation using copper oxide varistors. The direction of the "arrow" indicating the varistor is the direction of low resistance. Thus, when the carrier voltage varies in polarity, the equivalent circuits are as shown by (*b*) and (*c*).

Modulation and demodulation are the same fundamental processes of distortion by virtue of which the desired new signals (sidebands) are created. In modulation, the information contained in the voice-frequency wave is translated (or moved) to a new and higher frequency band (the sidebands) for transmission. In demodulation, the information contained in the received sideband is translated back to the original voice-frequency band to operate the telephone receiver of the listener. Thus, if the received sideband extends from 6500 to 9800 cycles and the locally generated carrier is 10,000 cycles, the $f_c - f_v$ term (page 420) will be an audible speech signal occupying a band of frequencies from 200 to 3500 cycles.

Improved Type C Carrier Telephone System.[31, 32, 33] In about 1938 an improved type C system was developed. The improved system incorporated changes that were revolutionary, particularly with regard to modulation and demodulation. The improved system provides three talking channels in about the same frequency range (about 6000 to 30,000 cycles) as the original system. It is, however, smaller and cheaper and gives improved transmission performance. Amplifiers[31, 34] with negative feedback (page 300) are used in the carrier repeaters which are spaced at intervals of 125 to 250 miles on open-wire lines. Regulating circuits adjust for changes in line attenuation.

The method of selecting the various sidebands and other details are much as shown in Fig. 20.

Modulation in Improved Type C Systems. Copper oxide varistors are used for modulation in the improved system instead of vacuum tubes as in the original system. Many circuit arrangements can be used;[35] the one being employed is shown in Fig. 24(a), and the shapes of the various waves are in Fig. 25.

The characteristics of a typical copper oxide varistor are shown in Fig. 43, page 310, and are similar to those shown in Fig. 4, page 274, for the selenium varistor. The numerical values of resistance for varistors of various types are different. In developing the theory of the varistor modulator the assumption is made that the carrier voltage is much larger than the modulating voltage and that the carrier voltage controls the circuit, giving it the characteristics shown in Figs. 24(b) and (c). This is because in the forward or conducting direction the resistance of a varistor unit approaches zero, but in the reverse or nonconducting direction the varistor resistance approaches infinity. Because of these variations, the current flowing through resistor R_2, which represents the impedance of the toll line leading to the distant receiving station, is as shown by Fig. 25(c). If this wave is examined critically, it is seen to contain the voice-frequency modulating signal and high-frequency components *other than the carrier,* because, theoretically, the varistor bridge circuit offers either zero or infinite impedance to the carrier voltage. Among the high-frequency components are upper and lower sidebands.

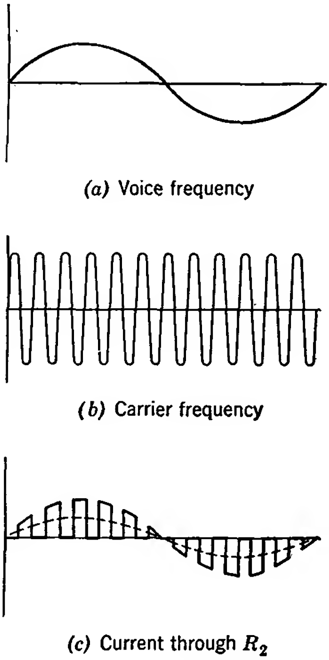


FIG. 25. If it is assumed that the carrier voltage of Fig. 24 is a control voltage, then the current that the audio voltage forces through R_2 of Fig. 24 is as shown by (c). This current contains sidebands and an audio (dotted) component.

The carrier voltage is assumed to control the switching of Fig. 24 and to have no other effect. The current through R_2 can be found by writing the equation for the **transfer impedance**[1] between the source of voice-frequency modulating voltage and R_2, and then dividing the

transfer impedance into the expression for the modulating voltage. Or the expression for the **transfer admittance** multiplied by the modulating voltage will give the modulated current through R_2.

The *transfer admittance* has the form of the wave composed of a constant term and alternating terms, because it is either zero as in Fig. 24(b) or $1/(R_1 + R_2)$ as in Fig. 24(c). The equation for a rectangular wave was given on page 318. Using this as a basis, and remembering that the transfer admittance is never negative, the equation for the instantaneous value of the transfer admittance becomes

$$y = \frac{1}{2(R_1+R_2)} + \frac{2}{\pi(R_1+R_2)}\left(\sin \omega_c t + \tfrac{1}{3}\sin 3\omega_c t + \tfrac{1}{5}\sin 5\omega_c t \cdots\right), \quad (10)$$

the values $\omega_c = 2\pi f_c$ being used because the transfer admittance is varying at the carrier rate. If the instantaneous modulating voice-frequency signal $E_v \sin \omega_v t$ is multiplied by the equation for the transfer admittance and if substitutions are made much as on page 420, it will be found that the desired sidebands have been created.

Demodulation in Improved Type C Systems. At the receiving end the carrier frequencies are separated from the voice frequencies on the same line by filters. The sidebands used for receiving are separated by filters from the sidebands used in transmitting. Next, each sideband is selected by band-pass filters for the proper channel. Demodulation is accomplished by a copper oxide varistor-bridge circuit essentially as in modulation. Because the carrier is suppressed at the transmitting end, a wave of the carrier frequency is generated locally and is injected into the demodulator circuit.

Type D Carrier Telephone System.[36] The type D carrier provides *one* talking channel and is sufficiently low in first cost and maintenance to be economical for installation on open-wire lines about 100 miles in length. It is a carrier-suppression system, transmitting one sideband. The frequencies used are indicated in Fig. 18. This system employs a self-oscillating modulator and demodulator, thus combining modulation and demodulation with carrier generation. The DA system contains an amplifier for increasing talking distances to about 200 miles.

Type G Carrier Telephone System.[37] The type G was the first carrier telephone system to use copper oxide modulators and demodulators; also, it uses *only one* vacuum tube. This is in the oscillator at the **active terminal.** No repeaters are used, the system being designed to give one carrier channel over circuits about 25 miles long.

Electric power is supplied by the oscillator at the active terminal.

The incoming voice-frequency signal wave and the carrier wave are impressed on a copper oxide varistor bridge to produce modulation. The carrier frequency *is transmitted* in this system, and hence the positions of the carrier frequency and voice frequency in Fig. 24 are interchanged.[38] Both sidebands are transmitted (Fig. 18). Demodulation occurs at the **inert terminal** in a copper oxide bridge circuit. When the person at the active terminal is *not* speaking, a carrier-frequency wave flows to the inert terminal so that, when the person at this terminal speaks, carrier is available for modulation. At *each* terminal the *same* copper oxide bridge circuit is used for modulation and demodulation and is sometimes called a **modem.**

Type H Carrier Telephone System.[39, 40] The type H is a single-channel system (Fig. 18) used on open-wire circuits about 100 to 200 miles long. Copper oxide modulators and demodulators are used, the carrier is suppressed, and one sideband is transmitted. This system is provided with a rectifier and can be operated from a 110-volt, 60-cycle source.

Multichannel Carrier Telephone Systems. For about 20 years the maximum carrier frequency used was approximately 30,000 cycles. Of course, higher frequencies *could* have been transmitted, but, because of high attenuation, crosstalk, and other factors, it was not commercially feasible to use higher frequencies.

About 1938 **multichannel carrier systems** were installed that utilized up to 140,000 cycles. Whereas the type C system gave only 3 talking channels, these so-called multichannel systems gave 12 talking channels for systems on conventional-type lines and cables. Later, 240 talking channels were provided on the original New York–Philadelphia coaxial cable, and *480* and then *600* talking channels were provided on coaxial cables in 1948.

In the design of these systems, much of the terminal equipment is standardized, so that the same type of apparatus may be used on the several systems. The voice-frequency band transmitted is from about 200 to 3500 cycles. The terminal equipment[41] employs copper oxide modulators and demodulators such as considered in the preceding pages, quartz crystal filters (page 187), and the magnetic generation of the various carrier frequencies.[41, 42]

This last development eliminates the necessity of many separate vacuum-tube oscillators in the following way. One vacuum-tube tuning-fork-driven oscillator supplies a single basic frequency to a non-linear network[42] which by distortion produces an output wave such as Fig. 26. This wave contains a fundamental and a large number of *odd* harmonics. The desired *even* harmonics are produced

by distorting the odd harmonics in full-wave copper oxide rectifiers. The separate harmonics are selected as desired by crystal filters and are impressed on the modulators and demodulators as carrier frequencies.

Type J Carrier Telephone System.[43, 44, 45] The type J carrier is designed *for open-wire lines* and operates in the region from about 36,000 to 140,000 cycles. Twelve talking channels are provided. A voice-frequency circuit and a type C carrier can operate simultaneously, giving a total of 16 channels over one pair of wires, as shown by Fig. 27. As Fig. 18 indicates, the band from 36,000 to 84,000 cycles is used for talking in one direction, and the band from 92,000 to 140,000 cycles is used for talking in the other direction.

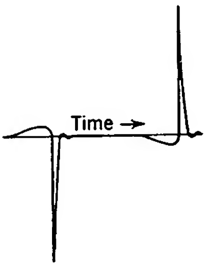


FIG. 26. Wave form of current in the output of the harmonic producer used to supply the carrier frequencies in multichannel carrier systems. (Reference 42.)

The incoming voice-frequency signals pass through hybrid coils (not shown in Fig. 27) and are there changed from two-wire to four-wire circuits as in any repeater or carrier circuit. Each of these voice-frequency signals is then impressed on one of the modulators at the top of Fig. 27. The carrier frequencies used with these modulators cover the band from 64 to 108 kilocycles. The lower sidebands of the outputs of these modulators are selected by the band filters and combined into a group. This group of from 60 to 108 kilocycles passes into the group modulator, which is supplied with a 340-kilocycle carrier frequency from an oscillator. Here modulation of the *group* occurs, and the band filter selects the group upper sideband that extends from 400 to 448 kilocycles. This group of frequencies is then impressed on a second group modulator together with a frequency of 484 kilocycles, if the system is for transmitting from *west to east.* The group lower sideband will be a group of frequencies extending from 36 to 84 kilocycles which is selected for transmission. This group contains the information of the twelve incoming voice-frequency channels. For *east to west* transmission, the last modulator uses a carrier frequency of 308 kilocycles, and the group of frequencies will be at 92 to 140 kilocycles for talking in the east-west direction. For receiving, the opposite sequence of translation is used.

Repeaters in Type J Carrier Systems.[43] Repeaters are required at 75- to 100-mile intervals, depending on wire sizes, etc. Even a very thin coating of ice on the line wires causes a great increase in the

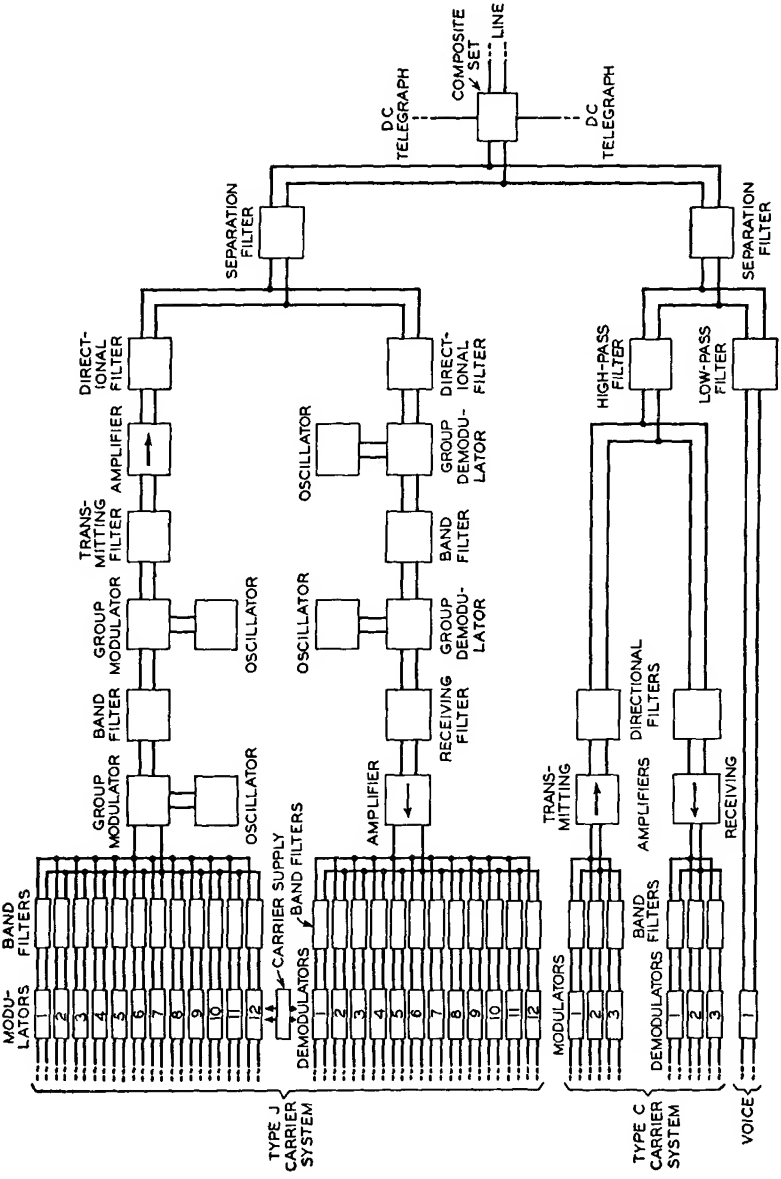


FIG. 27. The type J open-wire carrier telephone system provides twelve two-way channels on a single pair of wires that also provide three two-way type C carrier channels and one two-way voice circuit. (Courtesy *Bell System Technical Journal.*)

line attenuation, and, in areas where ice often forms, repeaters may be spaced about 50 miles apart. Pentode vacuum tubes are used in the amplifiers, which are of the negative-feedback type (page 300). Regulating amplifiers with pilot control are provided at intervals.

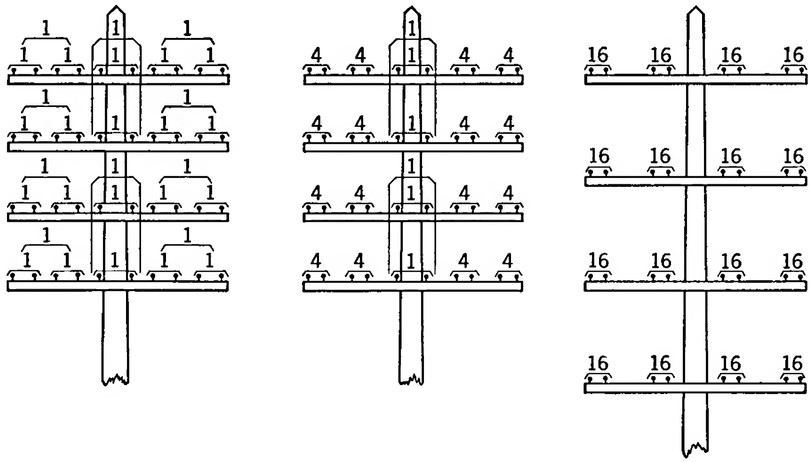


FIG. 28. At the left is shown the arrangement of crossarms and conductors for voice-frequency use only. Except for the pole pairs, 12-inch spacing is used between wires and 24-inch spacing between crossarms; including the phantoms, 30 two-way voice-frequency telephone circuits are provided.

In the center is shown a later construction (about 1925) with 8-inch spacing, except for pole pairs, and with 24-inch spacing between crossarms. One type C carrier system (3 channels) and one voice-frequency circuit operate on each of the 8-inch spaced pairs. Voice-frequency circuits operate on the pole pairs which are phantomed. With this arrangement 22 voice-frequency circuits and 48 carrier circuits are provided, giving a total of 70 two-way telephone channels.

At the right is shown a more recent construction (about 1940). The spacing of wires of a pair is 8 inches, the pole pairs are omitted, and the crossarms are spaced 36 inches apart. One voice-frequency circuit, one type C carrier (3 channels), and one type J carrier (12 channels) operate on each pair, giving 16 voice-frequency circuits, 240 carrier circuits, and a total of 256 two-way telephone channels. (Reference 43.)

Crosstalk in Type J Carrier Systems.[43, 44] Because of the serious crosstalk problem existing at the high frequencies used in this carrier system, a special transposition system (Chapter 14) was designed. Also, a new wire spacing of 8 inches between wires of a pair and 8 wires per crossarm was adopted (Fig. 28). Furthermore, transposition poles were more accurately located, and wire sags were reduced and equalized.

Type K Carrier Telephone System. The original type K1 multichannel carrier telephone system was *designed for non-loaded cables* and was placed in service about 1938. This provided 12 talking channels, and approximately 2,500,000 miles of two-way telephone message circuits were installed.[46] About 1944 an improved system, the type K2, was developed, and by 1947 an additional 2,500,000 circuit miles had been placed in service.[46]

Because the wires of a cable are so close together, there is a very great tendency for an exchange of energy to occur between circuits, resulting in serious crosstalk (page 565). This is true particularly at high frequencies, and in general the use of carrier in cables had been avoided prior to the development of the type K system. One factor that has made these systems possible is the use of *separate cables* for talking in the two directions. Or the same cable structure may be used, but the inner conductors are used for transmission in one direction, and the outer conductors used for transmission in the other direction, and a suitable shield is placed (during manufacture) between the two groups. In this way coupling between high-energy circuits, emerging from a repeater and low-energy circuits, just entering a repeater, is avoided.

Type K1 Carrier System. A block diagram of a type K1 system[41, 47] is shown in Fig. 29. Because separate cables are used for transmitting and receiving, the same numerical group of frequencies may be used for each direction, and hence the system is simplified as compared with the type J carrier. Briefly, the operation of the type K1 system is as follows: As indicated in Fig. 29, each of the 12 connected telephone sets feeds in through a hybrid coil (shown as a block) that separates the transmitting and receiving functions. The signal to be transmitted is impressed on a copper oxide modulator. The various carrier frequencies (obtained as explained for the type J system) are shown by arrows in Fig. 30. They are made high so that crystal filters can be used, and for other reasons.

After modulation, the lower sidebands are combined and fed in common to a group modulator together with a carrier frequency of 120 kilocycles. The lower sideband of this group is selected for transmission. Thus, the various telephone conversations are transmitted over the cable within the band 12,000 to 60,000 cycles. At the receiving end they are restored to their original audible frequencies by a similar double-demodulating process.

Repeaters in Type K1 Carrier Systems.[47] As indicated in Fig. 29, repeaters are placed in the cable circuits at about 17-mile intervals. Because separate cable pairs are used for transmitting in the two direc-

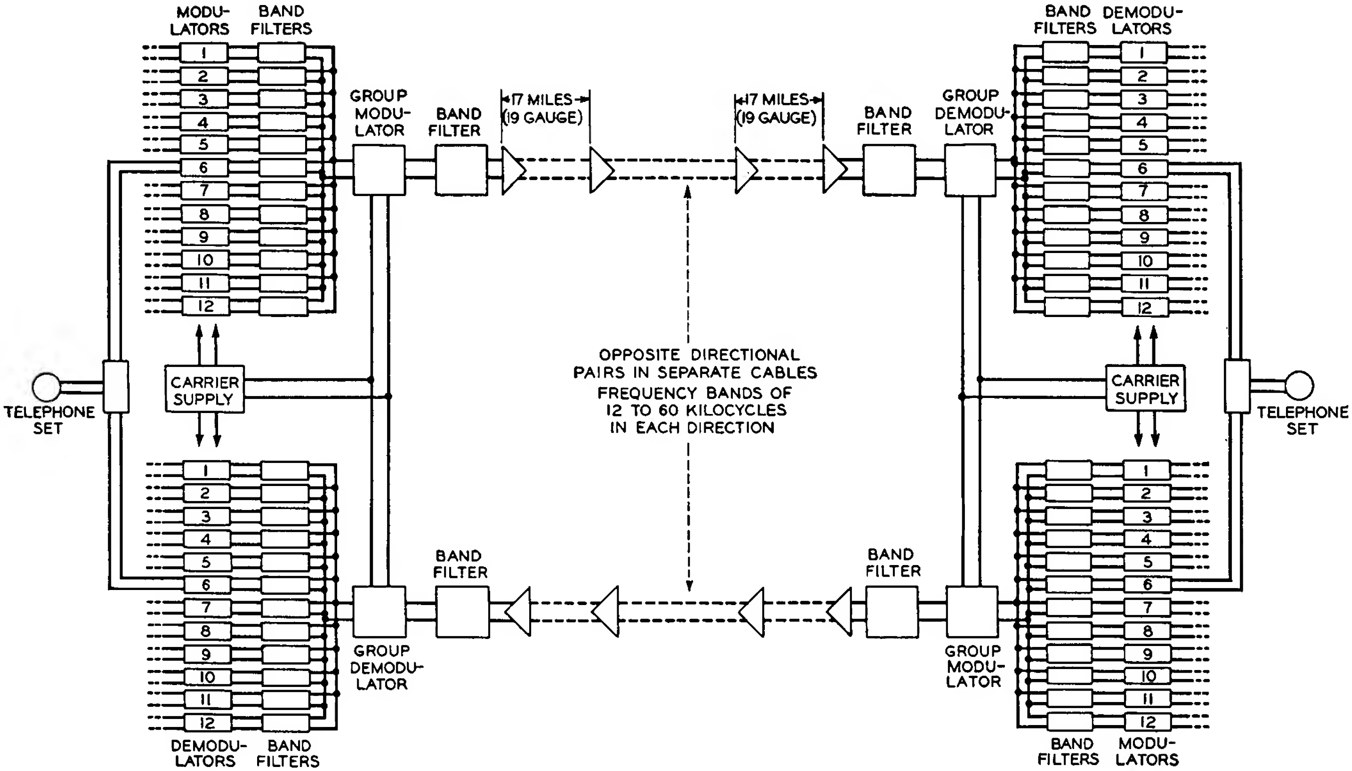


FIG. 29. Schematic of the type K cable carrier system, which provides 12 channels in each direction. The telephone sets are connected directly to the hybrid coil which separates the transmitting and receiving paths. The triangular blocks at 17-mile intervals are repeaters. (Courtesy *Bell System Technical Journal.*)

tions, at a repeater point two one-way vacuum-tube amplifiers are installed. As previously stated, the cable pairs used for type K systems are non-loaded, the reason being evident in Fig. 31. The wave velocity is about 100,000 miles per second which minimizes echo effects and reduces difficulties in conversation.[47, 48]

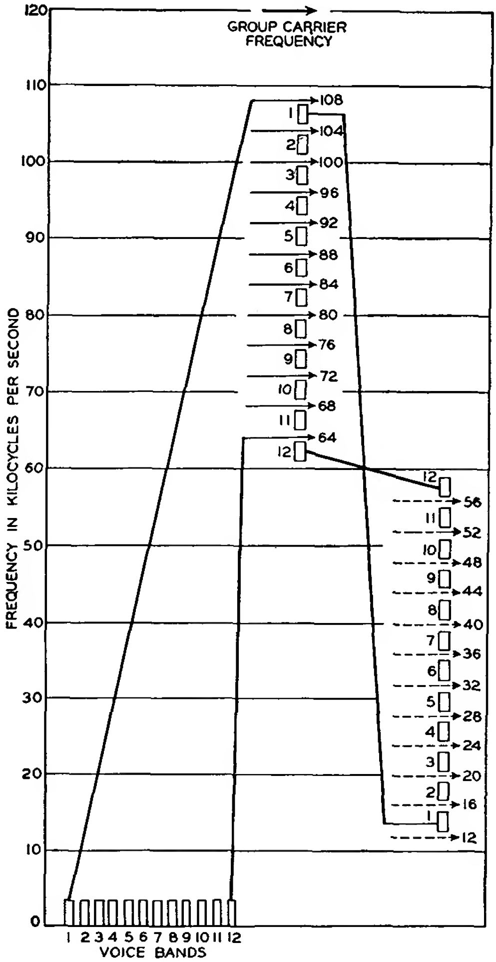


FIG. 30. By double modulation, twelve voice-frequency channels are first raised to occupy the band from 60 to 108 kilocycles, and then lowered to the band from 12 to 60 kilocycles for transmission over the cable. (Reference 41.)

As Fig. 31 indicates, temperature changes cause decided variations in cable attenuation. Temperature variations of underground cable occur slowly and ordinarily are not extreme, but for aerial cables the opposite is true. When the sun is shining directly on a lead-covered aerial cable the cable temperature may rise from 15° to 25° F. above that of the surrounding air; then, within a few minutes, the cable may be subjected to cold rain, following a thunder storm. The attenuation change with temperature variations for a cable is composed of two parts: *first*, a component that is *independent* of frequency, and *second*, a component that *varies* with frequency and is called[47] **twist.** The pilot-wire method (page 408) and automatic repeater gain adjustments are used to maintain the transmission at the correct level.

Crosstalk and Noise in Type K1 Carrier Systems.[47, 49] As has been explained, to reduce crosstalk, circuits in different cables and shielded circuits have been used for transmitting and receiving. This

minimizes what is called **near-end crosstalk**[1] (page 565) in the cables, and special precautions are taken to minimize crosstalk in office wiring. **Far-end crosstalk**[1] (page 566) between adjacent pairs transmitting in the same direction is reduced by connecting small adjustable mutual inductance **balancing coils** between the circuits and by other means. Noise originating in repeater stations is reduced by filters.

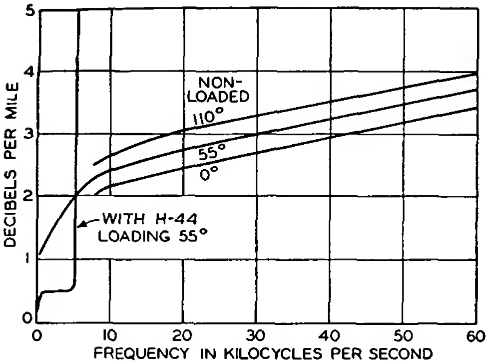


FIG. 31. Attenuation characteristics for loaded and non-loaded 19-gauge cable pairs. Temperature in degrees Fahrenheit. (Courtesy *Bell Laboratories Record.*)

Improved Type K2 Carrier System.[46] The basic circuit arrangement is the same as described for the type K1 carrier. The K2 carrier is particularly suited for transcontinental cable operation, gives improved performance, and is more economical than the older type. The most important changes are in the repeaters. Three types of amplifiers are used: **transmitting amplifiers** that are inserted only at the transmitting end; **twist amplifiers** that are placed about 100 miles apart on aerial cables, about 250 miles apart on underground cables, and compensate for the twist previously discussed; and **line amplifiers** that are placed approximately 17 miles apart. Line amplifiers do three things:[46] *first,* they compensate for the loss in the preceding cable section; *second,* they equalize for variations in loss at the different frequencies; and *third,* they regulate the gain automatically.

Feedback amplifiers with thermistor (page 310) control to adjust the gain are employed.[44]

Type L Coaxial Cable Carrier Telephone Systems. The United States is spanned with a coaxial cable circuit which was completed about 1948. This circuit can be used for at least two purposes: *first,* for the transmission of television programs (page 533); and *second,* for providing a large number of talking channels. Television requires a band of frequencies at least 4,500,000 cycles wide. About 480 telephone talking channels are provided over one coaxial system, and this number has been increased to 600 channels in some instances. Coaxial cables have been used experimentally for television purposes for some years.[50]

The type L carrier system (and modifications such as the type L1 system) provide telephone channels over coaxial cables. A coaxial cable has no cutoff frequency; however, at high frequencies the loss becomes so great that the repeaters must be placed very close together. There is, therefore, always an upper frequency range above which it is uneconomical to transmit with the equipment and methods available.

As previously mentioned, much of the terminal equipment for the types J, K, and L carrier systems is the same. Thus, a complete discussion of the type L system for coaxials is not necessary. The

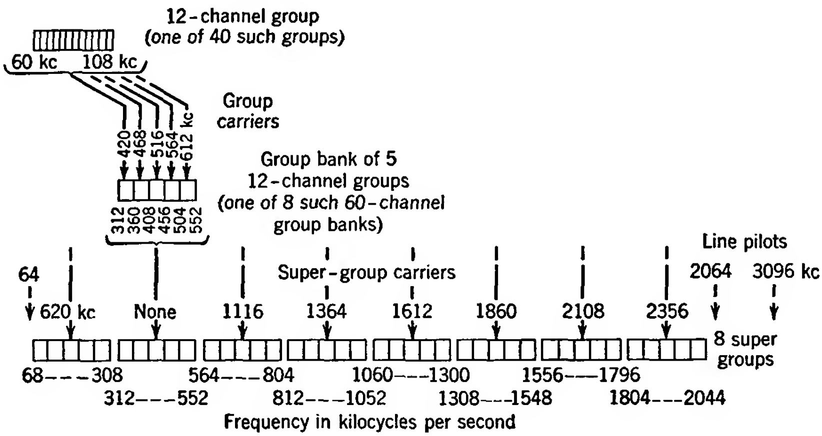


FIG. 32. In the type L1 carrier system used on coaxial telephone cables there are three amplitude-modulation steps used to translate, or move, the 480 (or more) telephone conversations to their respective assigned frequency bands. (Reference 51.)

method[51] used in arranging 480 telephone conversations for transmission over a coaxial system is as follows: The 480 incoming subscriber telephone lines are separated into groups of 12 and are impressed on carrier equipment such as is discussed in the preceding pages. Each group of 12 channels emerges from the modulators in the frequency band of 60 to 108 kilocycles. Five of these 60- to 108-kilocycle bands (each of which contains 12 conversations) are separated into groups and modulated with "group carriers" so that the 60 telephone conversations they contain lie in the 312- to 552-kilocycle band. Each of these bands (with the exception of one, see Fig. 32) is then modulated with "super-group carriers" and placed in the assigned region in the frequency spectrum. A comparable process is used at the receiving end.

One coaxial tube is used for transmission in one direction, and another tube for transmission in the opposite direction. As many as eight of these tubes are combined into a single sheath, providing four coaxial systems.[52] The early coaxial cables used a 13-gauge copper center conductor and a 0.27-inch copper outer conductor. For 480 telephone channels repeaters were necessary at approximately 5.4-mile intervals. A later coaxial cable has an outer diameter of 0.375 inch and repeaters every 7.9 miles. The repeaters use vacuum-tube amplifiers[50] specially designed to pass the entire frequency band transmitted.

Miscellaneous Carrier Systems. Carrier telegraph systems are used as discussed in Chapter 9. In addition to the applications described in the present chapter, carrier telephone systems are used on power lines for supplying telephone service to rural customers,[53, 54] for operational communication (and other purposes, such as relaying) on power lines,[55] and for communication on submarine cables.[56]

The rural-line carrier telephone system just mentioned is designated type M1. Although it is designed primarily for operation over rural electric lines as stated, it is also sometimes installed on a pair of telephone line wires to provide an additional talking channel. In such instances frequencies as high as 450,000 cycles are transmitted over ordinary telephone lines.

REFERENCES

1. *American Standard Definitions of Electrical Terms.* A.I.E.E., 1941.
2. Technical Developments Underlying the Toll Services of The Bell System. *Supplement to Bell System Tech. J.*, July, 1936.
3. Osborne, H. S. A General Switching Plan for Telephone Toll Service. *Bell System Tech. J.*, July, 1930, Vol. 9, No. 3.
4. Crossbar Toll. A series of articles reprinted from the *Bell Lab. Record,* Volumes 22 and 23.
5. Abraham, L. G., A. J. Busch, and F. F. Shipley. Crossbar Toll Switching System. *Elec. Eng.*, June, 1944, Vol. 63, No. 6.
6. Pilliod, J. J., and H. L. Ryan. Operator Toll Dialing—A New Long Distance Method. *Bell Telephone Magazine,* 1945, Vol. 24, No. 2.
7. Hatton, W. Automatic Ticketing of Long-Distance Telephone Connections. *Elec. Commun.*, January, 1940, Vol. 18, No. 3.
8. Hatton, W. Automatic Ticketing of Telephone Calls. *Elec. Commun.,* 1944, Vol. 21, No. 4.
9. Friend, O. A. Automatic Ticketing of Telephone Calls. *Elec. Eng.*, March, 1944, Vol. 63, No. 3.
10. Friend, O. A. Automatic Ticketing. *Bell Lab. Record,* July, 1944, Vol. 22, No. 11.
11. Shipley, F. F. Nation-wide Dialing. *Bell Lab. Record,* October, 1945, Vol. 23, No. 10.

12. Gherardi, B., and F. B. Jewett. Telephone Repeaters. *Trans. A.I.E.E.*, November, 1919, Vol. 38.
13. Miller, K. B. *American Telephone Practice*. . McGraw-Hill Book Co.
14. McIlwain, K., and J. G. Brainerd. *High Frequency Alternating Currents.* Second Edition. John Wiley & Sons.
15. Case, R. L. Repeaters for Two-Wire Toll Circuits. *Bell Lab. Record,* August, 1931, Vol. 9, No. 12.
16. Crawford, C. G. Four-Wire Telephone Circuits. *Bell Lab. Record,* September, 1931, Vol. 10, No. 1.
17. Edwards, P. G. V1 Telephone Repeater Arrangements. *Bell Lab. Record,* September, 1941, Vol. 20, No. 1.
18. Nance, H. H., and O. B. Jacobs. Transmission Features of Transcontinental Telephony. *Trans. A.I.E.E.,* Vol. 45, 1926.
19. Johnson, E. D. Transmission Regulating System for Toll Cables. *Bell Lab. Record,* January, 1929, Vol. 7, No. 5.
20. Clark, A. B. Telephone Transmission over Cable Circuits. *Trans. A.I.E.E.,* Vol. 42, 1923.
21. Clark, A. B., and H. S. Osborne. Long Distance Telephone Circuits in Cable. *Bell System Tech. J.,* October, 1932, Vol. 11, No. 4.
22. Clark, A. B., and R. C. Mathes. Echo Suppressors for Long Telephone Circuits. *Trans. A.I.E.E.,* Vol. 44, 1925.
23. Shea, T. E. *Transmission Networks and Wave Filters.* D. Van Nostrand Co.
24. Johnson, K. S. *Transmission Circuits for Telephonic Communication.* D. Van Nostrand Co.
25. Clark, A. B., and C. W. Greene. Long Distance Cable Circuits for Program Transmission. *Trans. A.I.E.E.,* October, 1930, Vol. 49, No. 4.
26. Lane, C. E. Phase Distortion in Telephone Apparatus. *Bell System Tech. J.,* July, 1930, Vol. 9, No. 3.
27. Colpitts, E. H., and O. B. Blackwell. Carrier Current Telephony and Telegraphy. *Trans. A.I.E.E.,* Vol. 40, 1921.
28. Hartley, R. V. L. Relations of Carrier and Sidebands in Radio Transmission. *Bell System Tech. J.,* April, 1923, Vol. 2, No. 2.
29. Bower, M. M. The Carrier Telephone Alphabet. *Bell Lab. Record,* February, 1938, Vol. 16, No. 6.
30. Affel, H. A., C. S. Demarest, and C. W. Green. Carrier Systems on Long Distance Telephone Lines. *Bell System Tech. J.,* July, 1928, Vol. 7, No. 3.
31. O'Leary, J. T., E. C. Blessing, and J. W. Beyer. An Improved Three-channel Carrier Telephone System. *Bell System Tech. J.,* January, 1939, Vol. 18, No. 1.
32. Montfort, L. R. The C5 Carrier System. *Bell Lab. Record,* August, 1940, Vol. 18, No. 12.
33. Cowley, G. W. The C5 Carrier Terminal. *Bell Lab. Record,* October, 1940, Vol. 19, No. 2.
34. Perkins, E. H. Repeaters for the C5 Carrier System. *Bell Lab. Record,* January, 1941, Vol. 19, No. 5.
35. Caruthers, R. S. Copper Oxide Modulators in Carrier Telephone Systems. *Bell System Tech. J.,* April, 1939, Vol. 18, No. 2.
36. Black, H. S., M. L. Almquist, and L. M. Ilgenfritz. Carrier Telephone System for Short Toll Circuits. *Trans. A.I.E.E.,* January, 1929, Vol. 48.

37. Jordan, H. G. New Carrier System Type G-1. *Bell Lab. Record,* August, 1936, Vol. 14, No. 12.
38. Blessing, E. C. Modulating in the G-1 Carrier System. *Bell Lab. Record,* March, 1937, Vol. 15, No. 7.
39. Dickieson, A. C. The Type-H Carrier Telephone System. *Bell. Lab. Record,* November, 1937, Vol. 16, No. 3.
40. Fisher, H. J., M. L. Almquist, and R. H. Mills. A New Single Channel Carrier Telephone System. *Bell System Tech. J.,* January, 1938, Vol. 17, No. 1.
41. Chesnut, R. W., L. M. Ilgenfritz, and A. Kenner. Cable Carrier Telephone Terminals. *Elec. Eng.,* May, 1938, Vol. 57, No. 5.
42. Peterson, E., J. M. Manley, and L. R. Wrathall. Magnetic Generation of a Group of Harmonics. *Elec. Eng.,* August, 1937, Vol. 16, No. 8.
43. Kendall, B. W., and H. A. Affel. A Twelve-channel Carrier Telephone System for Open-wire Lines. *Elec. Eng.,* July, 1939, Vol. 58, No. 7.
44. Iglenfritz, L. M., R. M. Hunter, and A. L. Whitman. Line Problems in the Development of the 12-Channel Open-wire Carrier System. *Elec. Eng.,* December, 1939, Vol. 58, No. 12.
45. Starbird, L. C., and J. D. Mathis. Some Applications of the Type J Carrier System. *Elec. Eng.,* December, 1939, Vol. 58, No. 12.
46. Black, H. S., F. A. Brooks, A. J. Wier, and I. G. Wilson. An Improved Cable Carrier System. *Trans. A.I.E.E.,* 1947, Vol. 66.
47. Green, C. W., and E. I. Green. A Carrier Telephone System for Toll Cables. *Elec. Eng.,* May, 1938, Vol. 57, No. 5.
48. Blackwell, O. B. The Time Factor in Telephone Transmission. *Trans. A.I.E.E.,* 1932, Vol. 51.
49. Weaver, M. A., R. S. Tucker, and P. S. Darnell. Crosstalk and Noise Features of Cable Carrier Telephone System. *Bell System Tech. J.,* January, 1938, Vol. 17, No. 1.
50. Strieby, M. E. Television Transmission by Coaxial Cable. *Elec. Eng.,* June, 1938, Vol. 57, No. 6.
51. Crane, R. E. Terminal Equipment for the L1 Carrier. *Bell Lab. Record,* December, 1941, Vol. 20, No. 4.
52. Abrahams, L. G. The Coaxial Conductor Expands in Size. *Bell Lab. Record,* November, 1946, Vol. 24, No. 11.
53. Barstow, J. M. Carrier Telephones for Farms. *Bell Lab. Record,* October, 1947, Vol. 25, No. 10.
54. Edson, R. C. M1 Carrier: The Common Terminal. *Bell Lab. Record,* February, 1948, Vol. 26, No. 2.
55. Cheek, R. C. A Comparison of Amplitude-modulation, Frequency-modulation, and Single-sideband Systems for Power-line Carrier Transmission. *Elec. Eng.,* May, 1945, Vol. 64, No. 5.
56. Affel, H. A., W. S. Gorton, and R. W. Chesnut. A New Key West–Havana Carrier Telephone Cable. *Bell System Tech. J.,* April, 1932, Vol. 11, No. 2.

REVIEW QUESTIONS

1. What distinction may be drawn between toll calls and long distance calls?
2. What is meant by the term telephone channel?
3. What is the distinction between toll trunking method and toll operating method?

4. How have toll calls been completed in past years, and what major change will probably be made in the future?
5. Discuss the numbering system to be used for nationwide dialing.
6. Explain why it is necessary to have repeaters at intervals along a transmission circuit, rather than merely at the terminals.
7. Explain the basic principle of the Shreeve telephone repeater.
8. What network theorem is involved in the changeover from Fig. 5(a) to 5(b)?
9. Name several reasons why the 22-type repeater is superior to the 21-type.
10. Consider the chapter as a whole, and enumerate reasons why four-wire telephone circuits are superior to two-wire circuits. What is an important disadvantage?
11. Describe the nature and causes of echoes on telephone systems.
12. Would balancing networks such as are shown in Fig. 14 be satisfactory for use with loaded cables? Why?
13. What type of modulation is used in carrier telephone systems?
14. In carrier telephony why is only one sideband transmitted? Would a system operate satisfactorily if both sidebands were transmitted and the carrier were supplied at the receiving end?
15. On page 413 it is stated that in amplitude modulation, sum and difference frequencies are created. Apply this principle to a device such as an audio-frequency transformer.
16. Explain how the carrier "carries" the information to the distant station.
17. What reasons are there for considering modulation and demodulation as the same fundamental process?
18. Explain how the carrier-frequency component is canceled in a push-pull modulator.
19. Enumerate reasons why copper oxide varistor modulators are used.
20. Approximately how many telegraph circuits could operate over one type C carrier system? Over one type J carrier system? Explain how this could be accomplished.
21. What novel features are incorporated in the type G carrier telephone system?
22. Enumerate some of the developments that have made possible the types J and K systems.
23. On page 426 it is stated, "This group of from 60 to 108 kilocycles, etc." Explain why the group has these frequencies.
24. As mentioned on page 426, ice on telephone line wires causes an increase in attenuation. What type of loss is this?
25. Several of the carrier systems use pilot control. Explain fully what this means; explain also the principle of operation.
26. Give two important reasons why cables for type K carrier are not loaded.
27. How would a shield be made so that part of a cable could be used for carrier transmission in one direction, and the rest for carrier transmission in the other direction? Can you explain how an ordinary telephone cable could be arranged so that it could be operated with some degree of satisfaction for transmission in opposite directions?
28. What is meant by the double-demodulating process mentioned on page 429?
29. Why may twist amplifiers on underground cables be farther apart than on aerial cables?
30. Is it probable that more than 600 telephone channels may be provided over coaxial cables? Give arguments for and against.

PROBLEMS

1. In Fig. 11 it is shown that about 65 microwatts are delivered to New York. If no intermediate repeaters are used, calculate the power that must be delivered to the line at San Francisco to supply this power to New York. Compare this with the total amount of power added along the line as shown in Fig. 11. If the characteristic impedance is 600-ohms resistance, calculate the input voltage and current for the first condition.
2. Assume that a 19-gauge cable with H-44-S loading extends from San Francisco to New York. If no repeaters are used, what power must be delivered to the cable at San Francisco to deliver 65 milliwatts to New York? If repeaters are used every 50 miles, what total power must be added to deliver 65 microwatts at New York? Assume that the input impedance of the loaded cable is about 820 ohms pure resistance, and calculate the input voltage and current for the first condition.
3. A non-pole pair side circuit of a 165-mil open-wire line is 1000 miles long. If the average daily attenuation change due to temperature variations is 5 per cent, what is the change in decibels? Calculate the corresponding change in decibels for a 19-gauge aerial cable with H-44-S loading. Of what practical significance are these figures?
4. On page 421 it is stated that the balanced modulator eliminates the carrier. Prove this to be true mathematically.
5. Make the substitution suggested on page 424 to prove that sidebands are created in the copper oxide bridge modulator.
6. Use a method similar to that employed in Fig. 24 to determine the output current wave shape when the carrier and sidebands are transmitted as in the type G carrier telephone system.
7. Calculate the characteristic impedance of the early coaxial cable. Do the dimensions used give the correct ratio (page 434) for minimum attenuation? If this same ratio is used for the 0.375-inch coaxial cable, what will be the nearest wire size used for the center conductor? Is this the size actually used?

CHAPTER 12

RADIO WAVE PROPAGATION AND ANTENNAS

Introduction. Wireless, or radio, communication has been used commercially since about 1900. This form of communication makes possible telephone and telegraph service between moving objects such as motor vehicles, trains, ships at sea, or aircraft in flight, and between locations separated by natural barriers such as oceans or extensive mountain ranges. This chapter and the following chapter will discuss radio communication.

Radio wave propagation is defined[1] as "the transfer of energy by electromagnetic radiation at radio frequencies." **Radio frequency** is defined[1] as a frequency "at which electromagnetic radiation of energy is useful for communication purposes."

Electromagnetic Waves. The energy used in radio communication is transmitted through space in the form of **electromagnetic waves,** defined[1] as a "wave in which there are both electric and magnetic displacements," or fields. The electromagnetic waves used in radio are the same, fundamentally, as those used in wire communication.

The energy in a cubic centimeter of space due to an electromagnetic wave can be shown[2] to be

$$W = \frac{1}{8\pi} (KE^2 + \mu H^2), \tag{1}$$

where K and μ are the dielectric constant and the magnetic permeability, respectively, of the transmitting medium (each unity for free space), and E and H are the electric and magnetic intensities of the wave.

The velocity of propagation of an electromagnetic wave in any medium can be shown[2] to be

$$V = \frac{c}{\sqrt{\mu K}}, \tag{2}$$

where c, the velocity of light in free space, has the value 3×10^{10}, and μ and K are as given in the preceding paragraph. In free space (approximately so for air), μ and K are both unity, and thus the

velocity of an electromagnetic wave is 3×10^{10} centimeters per second.

An electromagnetic wave has a **direction of polarization,**[1] which is the direction of the electric component or electric field. The discussions that follow will be limited to **vertically polarized waves** and **horizontally polarized waves.**

Production of Electromagnetic Waves. It is sometimes assumed that, when the voltage and current in a circuit produce electric and magnetic fields, *all* the energy entering the circuit is *stored* in these fields or is dissipated in heat. This is a reasonable assumption at low frequencies, but at *any* frequency *some* energy is radiated in electromagnetic waves, and the tendency for such radiation to occur varies as the *square* of the frequency.*

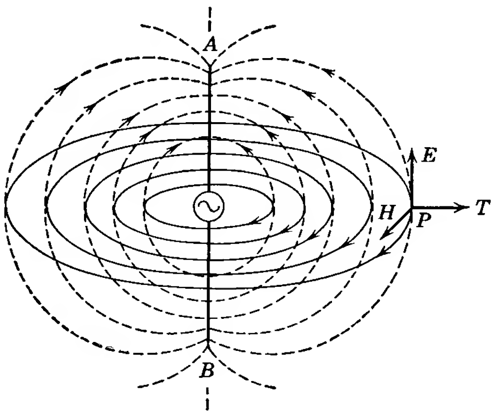


FIG. 1. Electric field (broken lines) and magnetic field around an antenna. The electric vector E and magnetic vector H are in *time* phase but are 90° apart in space. The direction of travel (propagation) is indicated by T. Electric and magnetic fields exist in planes other than the two planes shown.

Maxwell proved mathematically that, when the potential or current at a given point changes, the influence of this change is not felt at surrounding points immediately, but a definite time later, and that the propagation of electromagnetic waves was a possibility, a fact later established experimentally by Hertz.

The two wires forming the antenna of Fig. 1 are connected to an oscillator producing a radio-frequency voltage. The wires are about one-half wavelength long and are *in free space.* When the upper oscillator terminal is negative and the lower terminal is positive, electrons will flow up into wire A and up out of wire B. An electric current will flow down at this instant and will produce a magnetic field shown by the solid lines. Because wire A is negative and B is positive, a differ-

* A mathematical solution[3] gives the average power radiated by a short isolated wire to be $P_a = \omega^2 I^2 l^2/(3c^2)$, where $\omega^2 = (2\pi f)^2$, I is the current (maximum value), l is the length of the wire, and c is the velocity of light.

ence of potential will exist between these two wires, and an electric field will be established as shown by the broken lines. During the next half-cycle the applied radio-frequency voltage will reverse and the directions of the current flow, the potential differences, and the fields also will be reversed.

If point P of Figs. 1 and 2 is *very close* to the antenna, it is said to be in the **near zone,**[4] and in this region the **induction field** predominates. This is the field that is considered in low-frequency work to be directly proportional to the current and voltage. The *induction field* alternately accepts power from, and returns power to, the electric circuit. The intensity of the induction field in free space *varies inversely as the square of the distance* from the antenna.[2, 3]

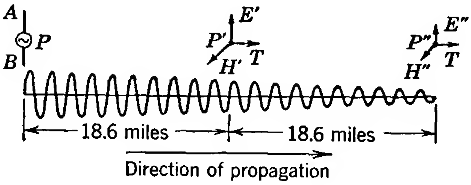


FIG. 2. A radio wave in space. The wave consists of electric (E) and magnetic (H) fields that vary sinusoidally. The decrease in intensity occurs because the wave "expands" from point source P. The frequency is 100,000 cycles.

At point P of Figs. 1 and 2 a **radiation field** exists in addition to the induction field described. In the near zone close to the antenna, the induction field is much stronger than the radiation field. At a distance from the antenna of approximately one-sixth of a wavelength,[3] the induction and radiation fields are of equal strength. Beyond this distance the **far zone**[4] exists, and the radiation field is the stronger. The intensity of the radiation field *varies inversely as the first power of the distance* from the antenna.[2, 3]

The production and propagation of an electromagnetic radio wave are illustrated in Fig. 2. The antenna A–B produces both induction and radiation fields at point P near the antenna. Beyond point P the induction field is negligible. The radiation field travels outward, and the electric and magnetic components reach point P' a finite time later than they were produced. Similarly, they reach point P'' at a later time. The lengths of the electric and magnetic vectors are less at P'' because the field intensity decreases inversely with the distance. In Fig. 2 E', E'' are the electric intensities, and H', H'' are the magnetic intensities of the radiation field. At a given point E and H are in *time* phase but are 90° apart in space. The fields at point P'' lag behind those at P', the amount depending on the distance. It is assumed in Fig. 2 that the frequency is 100,000 cycles and that the wave velocity is 186,000 miles per second.

The *induction field* is considered to be in a **quasi-stationary** state,[3] and the electric and magnetic components have no effect on each other. That is, in a given region there can exist a strong electric field and a weak magnetic field, and vice versa. Each field is in phase with the voltage or current producing it.

The *radiation field* is in a **dynamic state;**[3] a changing magnetic field has the ability to produce an electric field, and a changing electric field has the ability to produce a magnetic field. The radiation field is the portion of the total field about an antenna that, in a sense, cuts itself adrift from the total field produced. As an analogy, long after a stone has settled to the bottom of a pond, water waves travel over the surface of the pond.

Radio Frequencies Used in Communication. A classification for the high-frequency region of the radio-frequency spectrum was suggested[5] by the Bureau of Standards. To this classification the first three items have been added, giving Table I.

TABLE I

CLASSIFICATION OF RADIO FREQUENCIES

Name and Abbreviation	Frequency Kilocycles	Frequency Megacycles
Very low frequency, VLF	10 to 30	0.01 to 0.03
Low frequency, LF	30 to 300	0.03 to 0.3
Medium frequency, MF	300 to 3000	0.3 to 3
High frequency, HF	3000 to 30,000	3 to 30
Very high frequency, VHF	30,000 to 300,000	30 to 300
Ultrahigh frequency, UHF	300,000 to 3,000,000	300 to 3000
Superhigh frequency, SHF	3,000,000 to 30,000,000	3000 to 30,000

Name and Abbreviations	Wavelength Centimeters	Wavelength Meters
Very low frequency, VLF	3,000,000 to 1,000,000	30,000 to 10,000
Low frequency, LF	1,000,000 to 100,000	10,000 to 1000
Medium frequency, MF	100,000 to 10,000	1000 to 100
High frequency, HF	10,000 to 1000	100 to 10
Very high frequency, VHF	1000 to 100	10 to 1
Ultrahigh frequency, UHF	100 to 10	1 to 0.1
Superhigh frequency, SHF	10 to 1	.1 to 0.01

The Committee on Wave Propagation of the Institute of Radio Engineers has proposed the following frequency designations, in which cps is cycles per second, kc is kilocycles per second, mc is megacycles per second, and kmc is kilomegacycles per second.

TABLE II

FREQUENCY DESIGNATIONS

Frequency	Band Number
1 cps — 10 cps	0
10 cps — 100 cps	1
100 cps — 1000 cps	2
1 kc — 10 kc	3
10 kc — 100 kc	4
100 kc — 1000 kc	5
1 mc — 10 mc	6
10 mc — 100 mc	7
100 mc — 1000 mc	8
1 kmc — 10 kmc	9
10 kmc — 100 kmc	10

Radio Wave Transmission Paths. If a transmitting antenna were in *free space,* a radio wave would merely "spread out" as it traveled away from the antenna. But transmitting antennas are *not* in free space (even if they are some miles above the surface of the earth). Radio transmission is affected by the presence of the earth. Also, the upper atmosphere of the earth is a rarefied, ionized, conducting region that influences transmission considerably. This region is called the **ionosphere** and is defined[1] as "that part of the earth's atmosphere above the lowest level at which the ionization is large compared with that at the ground, so that it affects the transmission of radio waves." The lowest level of the ionosphere is about 50 kilometers (31 miles) above the surface of the earth.[1] Radio transmission also is influenced by the **troposphere** defined[1] as "that part of the earth's atmosphere in which the temperature decreases with altitude, clouds form, and convection is active." The troposphere extends for about 10 kilometers (6.2 miles) above the surface of the earth.[1]

Rays. In considering the transmission paths, it is convenient to assume that radio waves travel in rays. Of course, if a radio wave were visible, an oncoming radio wave would appear to be magnetic and electric lines of force crossed at right angles. This is called a **wave front.** It is much simpler to represent a radio wave by a **ray,** or line, perpendicular to the wave front, showing the direction of propagation of the wave. This has been done in Fig. 3.

Ionospheric Wave. If a transmitting antenna directs a component of its total radiated energy upward toward the ionosphere, under conditions often encountered, the ionosphere reflects much of this energy back toward the earth. This is indicated in Fig. 3. The **ionospheric wave** (or sky wave) is defined[1] as "a radio wave that is propagated by reflection from the ionosphere."

Tropospheric Wave. At very high radio frequencies and beyond, the troposphere commonly has much influence on radio transmission. A **tropospheric wave** is defined[1] as "a radio wave that is propagated from a place of abrupt change in the dielectric constant or its gradient with position in the troposphere."

Ground Wave. Particularly important in standard radio broadcasting is the **ground wave** defined[1] as "a radio wave that is propagated over the earth, and is ordinarily affected by the presence of the ground. The ground wave includes all components of a radio wave over the earth except ionospheric waves and tropospheric waves. The

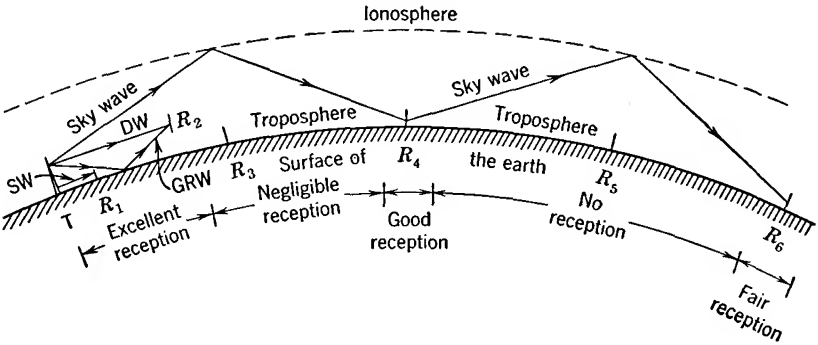


FIG. 3. This shows the various paths by which radio waves travel. This drawing is considerably simplified from the actual condition that may prevail at a given time. The transmitting antenna is marked *T*, and the receiving antennas at various points are R_1, R_2, R_3, etc. The abbreviations are DW, direct wave; GRW, ground-reflected wave; SW, surface wave. Reflections actually occur *in* the ionosphere and *in* the surface layers of the earth.

ground wave is somewhat refracted by the normal gradient of the dielectric constant of the lower atmosphere."

The ground wave is often considered[6] to be composed of two components, a **surface wave** and a **space wave.** The surface wave is the component of radiation that travels *entirely* along the surface of the earth (Fig. 3). The space wave is composed of a **direct wave** "that is propagated directly through space"[1] and a **ground-reflected wave,** which is "the component of the ground wave that is reflected from the ground."[1] These wave components or rays also are shown in Fig. 3.

Wave Diffraction. When a radio wave encounters an obstruction in its path, **diffraction** occurs. Thus, a large building or a hill will not cause a sharp "radio shadow," and reception may be possible behind a large building or below the top of a range of hills. Diffrac-

tion is one factor causing radio waves to bend with the surface of the earth.

Wave Refraction. When a radio wave passes from one medium having a certain dielectric constant and velocity of propagation to another medium having a different dielectric constant and velocity of propagation **refraction** occurs. Because of refraction, the path of a radio wave is bent, and, if the bending is sufficient, wave **reflection** may occur.

Absorption. If a radio wave were traveling in free space, a spreading would occur, yet there would be no energy dissipation in the usual sense. But a radio wave does not travel in free space. Thus, radio waves near the surface of the earth suffer an energy loss in this surface, and in buildings, trees, hills, etc. Also, a radio wave in the troposphere may suffer energy loss by absorption, probably caused by water vapor. Absorption of energy also occurs in the ionosphere.

The Troposphere.[3, 7] Energy absorption occurs in the troposphere. A second effect is that the air, being more dense at the surface of the earth, causes a radio wave, or ray, near the surface to travel more slowly than a ray in the upper troposphere, thus bending the waves around the earth. Hence, the direct and ground-reflected waves of Fig. 3 should be bent downward slightly. A third effect is explained as follows: From the definition on page 443 it would be concluded that an increase in altitude is *always* accompanied by a decrease in temperature. However, **temperature inversions** often occur in which moist warm air overlays cold dry air. Such discontinuities cause a wave to be refracted and may cause reflections to occur (page 446).

The Ionosphere.[8, 9] Previous to the transmission of the first wireless signals across the Atlantic in 1901, it was thought that such transmission would be impossible since electromagnetic waves traveled in straight lines. Kennelly and Heaviside independently and almost simultaneously suggested that such transmission was made possible by an upper ionized conducting region.

At the surface of the earth the atmosphere consists of about 76 per cent nitrogen, 23 per cent oxygen, and 1 per cent argon. At increased heights the oxygen and argon percentages decrease and the nitrogen increases until at about 60 miles the atmosphere is almost entirely nitrogen with but a small amount of oxygen and a trace of helium. After this point is passed the nitrogen decreases very rapidly until at about 100 miles the atmosphere is almost 95 per cent helium with but 5 per cent nitrogen.[10]

The air pressures at these heights are well described by the following quotation:[11]

> In the vicinity of 50 miles (80 kilometers) it has dropped to approximately 1/160,000 of that at the surface. This pressure is what we would ordinarily call a fair vacuum. At 125 miles (200 kilometers) the pressure is approximately 1/180,000,000 of an atmosphere, which is probably lower than the best vacuum usually obtained in the laboratory.*

The bending of the electromagnetic waves (rays) and their reflection back toward the earth by the ionosphere requires the presence of free electrons. These are produced by ionization of the rarefied gases just considered by radiations from various sources, chiefly ultraviolet light from the sun and cosmic rays from space, and by the action of electron streams shot out by the sun.[12]

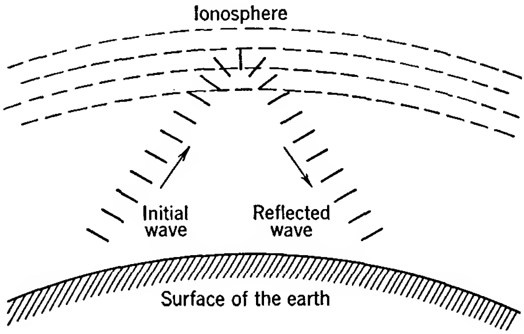


Fig. 4. The successive wave fronts of radio waves directed toward the ionosphere are shown by the short heavy lines. The upper portion of the wave front enters the ionosphere first and penetrates more deeply. The upper portion travels at a greater velocity, causing refraction, thereby turning the wave and reflecting it back toward the earth. Some energy absorption occurs in the ionosphere.

The refraction (Fig. 4) in the ionosphere that causes bending of the rays and their reflection back toward the earth is explained as follows:[12]

> When an electromagnetic wave traverses this reflecting region the unattached electrons are moved by the influence of the wave and absorb energy from it. On account of their mass and charge and the presence of the earth's magnetic field the electrons in their movements reradiate the absorbed energy slightly out of phase with the passing wave; in doing so they change the effective velocity within the refracting zone. This variation in effective velocity is, of course, the cause of refraction.†

* Reprinted by permission, courtesy C. W. Rice, and *QST*.

† Reprinted by permission, courtesy of R. A. Heising and the *Bell Laboratories Record*.

Ionospheric Regions. The ionosphere is considered as being composed of an ***E* region** between about 90 and 140 kilometers (56 and 87 miles) above the surface of the earth and an ***F* region** between 140 and 400 kilometers (87 and 250 miles) above the surface of the earth.[1]

Ionospheric Layers. Measurements of radio waves reflected back to the earth have identified the existence of several ionized layers in the regions of the ionosphere just described.

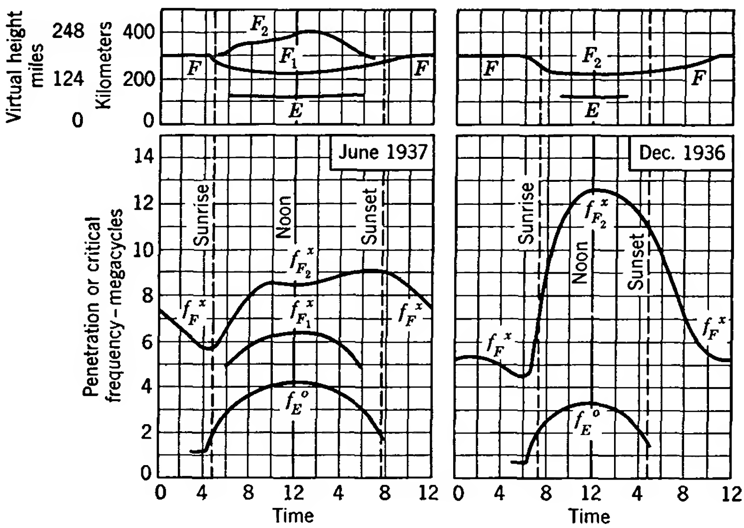


FIG. 5. Typical variations in the virtual heights of the ionosphere layers. The corresponding penetration, or critical, frequencies are shown. (Data from Reference 8.)

The *E layer*[1] is a "permanent" layer existing (Fig. 5) at about 110 kilometers (70 miles). Although it is permanent, in that it is present day after day, the ionic density of the *E* layer is *not* constant. It is strongest during the day, and at night it may almost disappear; also, it is subject to erratic variations.

The *F layer*[1] is a "permanent" layer at about 300 kilometers (185 miles) that varies as shown in Fig. 5. During all except the winter months, the *F* layer divides in the daytime into the F_1 and F_2 layers. The *F* layer also is subject to erratic variations.

A *D layer* has been found occasionally in the daytime at 50 to 90 kilometers (30 to 56 miles), but relatively it is of little importance.

In addition to the daily and seasonal variations depicted in Fig. 5, the ionosphere and its "layers" are affected by the 11-year sunspot

cycle and by the number of sunspots present.[13] The ionosphere is subject to sudden changes during erratic solar activity and the related phenomena of the aurora borealis and magnetic storms.[14] The word *layers* was placed in quotation marks because they do not exist as layers in the usual sense. This is indicated at the right in Fig. 6 which shows a typical daytime distribution of ionization.[8]

Virtual Height. Assume that the ionosphere layers are as in the upper left portion of Fig. 5 and that at noon short pulses of radio-frequency energy are directed vertically upward toward the ionosphere.

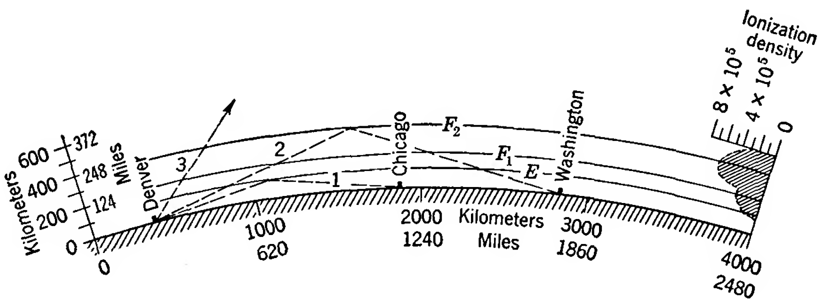


FIG. 6. The ionized layers of the ionosphere cause refraction and resulting reflection of radio waves as indicated for a typical daytime condition. The ionization density is shown at the right. Note that the several "layers" E, F_1, and F_2 are *not* distinct from each other and that the ionization density is greatest for the F_2 layer. The ionization density is measured in number of ions per cubic centimeter. (Adapted from Reference 8.)

If the frequency is below about 4.2 megacycles, some of the energy will be absorbed in the E layer, and the rest will be reflected back toward the earth. By properly displaying the initial pulse and the reflected pulse on a cathode-ray tube and by the use of suitable time-measuring equipment, the virtual height of the reflecting layer can be determined.[15] By **virtual height** is meant[1] the "height at which reflection from a definite boundary surface would cause the same time of travel as the actual reflection, for a wave transmitted from the ground to the ionosphere and reflected back." The virtual height is usually specified for the *lowest* frequency at which reflection occurs.[1] This is sometimes called the **equivalent height,** or the **effective height,** of the layer.

Penetration (Critical) Frequency.[1] Again assume that the ionosphere layers are as in the upper left portion of Fig. 5 and that at noon short pulses of radio-frequency energy are directed vertically upward toward the ionosphere. If the frequency is below about 4.2

megacycles for the period indicated, reflection back to the earth occurs, but, *above* this frequency (at noon), as indicated by the $f_E°$ *line, negligible reflection occurs from the E layer for waves directed vertically upward.* This frequency is defined as the **penetration frequency**[1] and is often called the **critical frequency.** It is the highest frequency reflected from a layer at vertical incidence.[1]

As mentioned, for the conditions depicted in the left portion of Fig. 5 at noon a wave of frequency greater than about 4.2 megacycles would not be reflected appreciably by the E layer. Some energy would be absorbed by the E layer, and the remainder of the wave would pass on into the F_1 layer, which is more dense (Fig. 6) and accordingly would reflect the wave. If, however, the frequency is too high for the F_1 layer, the wave might pass through to the F_2 layer, which is more dense (Fig. 6) and which might reflect the wave. Waves of frequencies greater than the penetration frequency for the F_2 layer are absorbed in part, and then transmitted out into space.

Because of the presence of the earth's magnetic field **double refraction** occurs in the ionosphere.[8] Hence, when a radio wave penetrates the ionosphere, two components emerge; these have different velocities, different polarization, and are attenuated differently. These two components are the **ordinary wave**[1] (or **ordinary ray**) and the **extraordinary wave**[1] (or **extraordinary ray**). The ordinary wave is designated by an "*o*" and the extraordinary wave by an "*x*". The ordinary wave is shown in Fig. 5 for the E layer because it is the more important. For the F layer the extraordinary wave is shown because, for many purposes, it alone is of importance.[8] A wave emerging from the ionosphere has both horizontally and vertically polarized components.

The penetration frequency is a measure of the ionization density, because, the higher the frequency, the greater must be the density of the ions to be able to reflect the wave back toward the earth. The relation applying is

$$N = 0.0124f^2,$$

where N is the number of ions per cubic centimeter and f is the penetration, or critical, frequency in kilocycles.[8]

Maximum Usable Frequencies. If a wave is directed vertically upward and if the frequency is sufficiently high, then no ionospheric reflection will occur. As previously explained, the penetration, or critical, frequency is the highest frequency reflected at vertical incidence, illustrated by ray A of Fig. 7. If the transmitting antenna is also sending (at this *same* frequency) ray B at the angle indicated,

some bending of the ray will occur, but not enough to cause reflection. If, however, a ray is sent at a **critical angle** θ, as with ray C, then the bending will be sufficient to cause reflection back to the earth. The **skip distance**[1] shown is "the minimum distance at which radio waves of a specified frequency can be transmitted at a specified time between two points on the earth by reflection from the regular ionized layers of the ionosphere. Reflected waves are received at less distance only by **sporadic, scattered,** or **zigzag reflections.**" These are erratic and unsatisfactory for reliable communication.

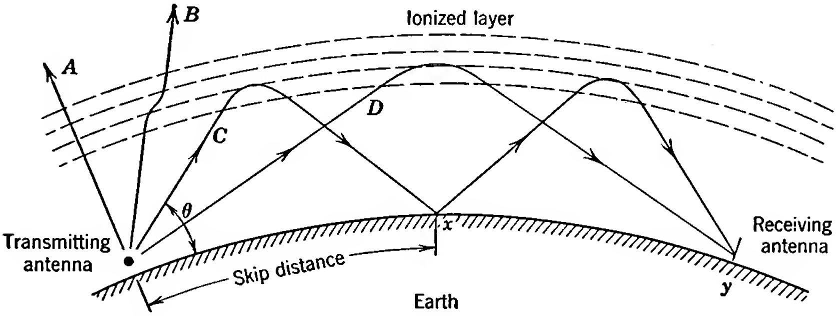


FIG. 7. Radio waves of the same frequency, but radiated at different angles, behave somewhat as shown. The angle θ is the critical angle for the frequency used.

When a ray such as C of Fig. 7 strikes the earth, absorption of energy occurs; also, the wave is reflected, and the signal may reach point y. A wave is said[1] to reach point x by a **single-hop path,** and to reach point y by a **double-hop path.** Ray D sent out by the antenna reaches point y by a single-hop path. The signal voltage induced in the receiving antenna will be due to the combined effects of the two (or more) rays.

The effect of the angle of radiation on radio signals of the *same frequency* was considered in the preceding paragraphs. The effect of frequency on radio signals radiated at the *same angle* will now be considered, using Fig. 8. At some frequency f_1 the ray will not be reflected but will pass even through the F layer. At a lower frequency f_2 the ray will pass through the E layer, but reflection back to the earth will occur from the F layer. At a lower frequency f_3 the ray will be reflected by the E layer.

From phenomena just considered, it follows that there is a **maximum usable frequency,** defined[1] as "the highest frequency that can be used for radio transmission at a specified time between two points on the earth by reflection from the regular ionized layers of the

ionosphere. Higher frequencies are transmitted only by sporadic and scattered reflections." For instance, suppose that the transmitting antenna of Fig. 7 is to provide a signal at point x. For a given ionosphere height (existing at that specific time), the signal can be provided by ray C which leaves the antenna at a critical angle for some

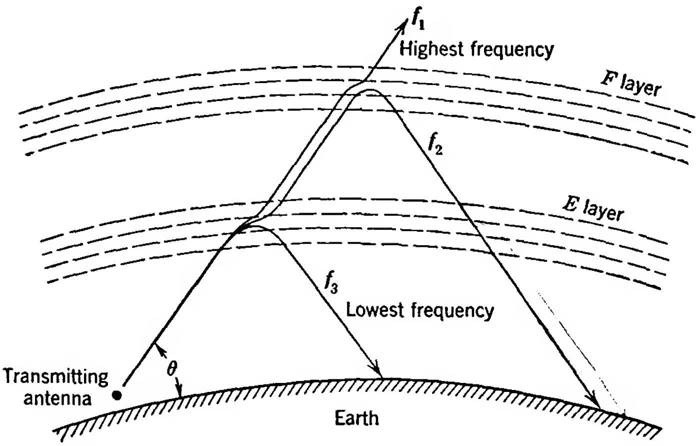


FIG. 8. Radio waves of different frequencies, but radiated at the same angle θ, behave somewhat as shown.

frequency. A higher frequency at this angle will not be reflected, and, if sent at a lower angle so that it will be reflected, it will miss point x. The maximum usable frequency f_m is (neglecting the earth's curvature) approximately[8]

$$f_m = f_c \sqrt{\frac{d^2}{4h^2} + 1}\,, \tag{3}$$

where d is the horizontal distance from the transmitting to the receiving antenna, h is the virtual height of the reflecting layer under consideration, and f_c is the penetration or critical frequency. The units d and h should be the same.

Radio Wave Reflections from the Earth. Ground-wave transmission is affected by the surface of the earth. This is also true for sky-wave transmission involving more than a single hop. The earth is not a perfect conductor.

Certain of the electrical constants[16] of the surface of the earth of importance in radio transmisison are as follows:

Type of Terrain	Dielectric Constant (Inductivity)	Conductivity e.m.u.	Absorption factor at 50 miles, 1000 kilocycles
Sea water	81	4.64×10^{-11}	1.0
Pastoral, rich soil, low hills, midwest	14	3×10^{-13}	0.50
Pastoral, medium soil, medium hills, east	13	6×10^{-14}	0.09
Rocky soil, steep hills, New England	14	2×10^{-14}	0.025
City, industrial area, average	5	1×10^{-14}	0.011

The dielectric constants given are with respect to air as unity. The conductivity values are in electromagnetic units and can be converted to mhos by multiplying by 10^9. The absorption factor is the ratio of the actual field intensity to the field intensity with no absorption[16] and indicates which terrain is the poorest, sea water being assumed perfect. These constants are determined by studying radio transmission, although it is possible to make direct measurements.[17]

If the ground were a *perfect conductor*, then the reflection of radio waves by the surface of the earth would be somewhat as explained in Chapter 6 for a transmission line. The electric field at the surface of a perfect conducting plane must be zero because there can be no voltage drop over such a surface. The reflected electric component must reverse to cause cancellation. From a study of these phenomena,[3] it is concluded, theoretically, that a horizontally polarized wave is reflected from a perfectly conducting surface with a 180° phase shift and that a vertically polarized wave is reflected without a phase shift.

Since the earth is not a perfect conductor, the theoretical situation just described may not exist. Studies of wave reflection *under actual conditions* have been made.[17] Both the ratio of the amplitude of the incident to reflected wave and the phase shift vary with the conductivity and the dielectric constant of the earth and with the frequency and angle of incidence. These values can be calculated[17] or taken from curves.[18]

Ground-Wave Propagation.[18, 19, 20, 21, 22] The Austin-Cohen formula[6] was used until about 1930 to calculate field strengths at a distance from a transmitting antenna. Its use has been largely superseded by the Sommerfeld formula as modified by several investigators.

Of particular interest is the problem of calculating the field strengths produced by standard amplitude-modulation broadcast stations. In engineering practice, curves[16] are usually used for this purpose; a summary of the method follows.

The field strength produced at a distance by a short vertical antenna

(about 0.1λ in length) over perfect earth is[23]

$$E_1' = 300\sqrt{P}, \tag{4}$$

where E_1' is the field strength in millivolts per meter at one kilometer when P is the power input to the antenna in kilowatts. For the field strength at one mile, the equation is[23]

$$E_1 = 186\sqrt{P}. \tag{5}$$

For an actual broadcast-type antenna, different factors[16] are used, and the equations are modified[23] as follows:

$$\text{Vertical antenna } 0.15\lambda \text{ to } 0.25\lambda \text{ high, } E_1 = 150\sqrt{P}. \tag{6}$$

$$\text{Vertical antenna } 0.25\lambda \text{ to } 0.40\lambda \text{ high, } E_1 = 175\sqrt{P}. \tag{7}$$

$$\text{Vertical antenna } 0.40\lambda \text{ to } 0.60\lambda \text{ high, } E_1 = 220\sqrt{P}. \tag{8}$$

In these three equations, E_1 is the field strength in millivolts per meter *at one mile* and P is the *power input* to the antenna in kilowatts. The constants for these three equations include such factors as the effect of height and the efficiencies of the antenna and ground system. Note that the field strength is proportional to the square root of the power input.

The strength of the ground wave at distances greater than one mile is found from curves. These are available[6, 16, 21, 22, 23] for various earth conductivities and dielectric constants. In Fig. 9 are included[22] curves for good earth conditions, such as encountered in midwestern United States. These curves are based on a field strength of 186 millivolts per meter at one mile and, for a station that does not produce this signal strength, the ordinates of Fig. 9 should be multiplied by the ratio of the actual field at one mile to 186. These curves indicate that ground-wave propagation at high frequencies becomes very inefficient because of losses in the surface of the earth and in objects on the surface. For this reason, high-frequency transmission is not used where ground-wave reception is desired. An inverse-distance curve also is given in Fig. 9. This represents transmission over lossless earth, and the decrease is caused by **dispersion,** or spreading out of the wave.

Signal strengths at a receiving location for good reception are as follows:[16]

City business or factory areas	10 to 50 millivolts per meter.
City residential areas	2 to 10 millivolts per meter.
Rural areas	0.1 to 1.0 millivolt per meter.

Radio reception is entirely satisfactory under favorable static and noise conditions with weaker signals, perhaps as low as 10 microvolts per meter and even less. Theoretically, 1 **millivolt per meter** will induce a voltage of 1 millivolt in a wire one meter long if the wire is

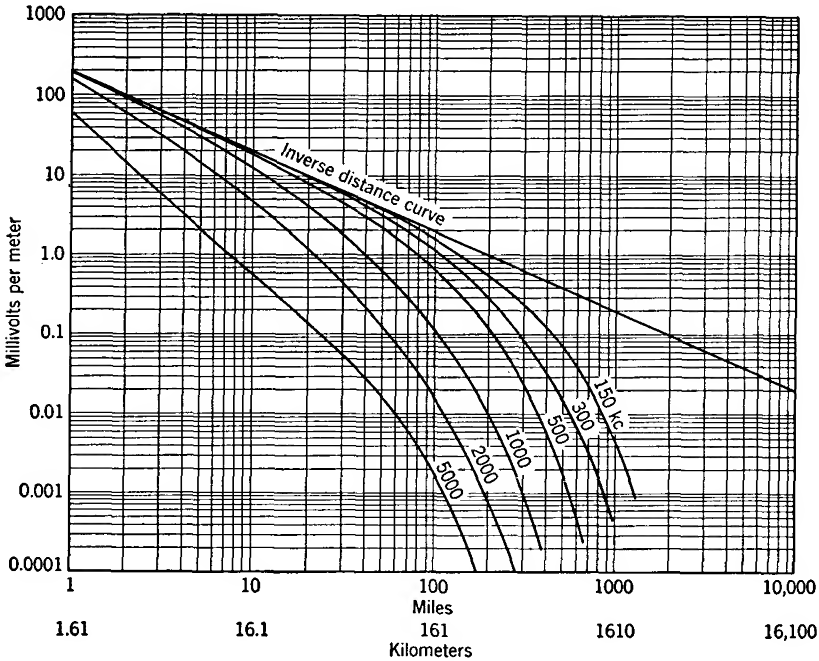


FIG. 9. Curves for computing ground-wave signal intensity for good earth conditions (conductivity $\sigma = 10^{-13}$ e.m.u., dielectric constant $\epsilon = 15$). The inverse distance curve is the curve for perfect earth where there is no loss and the signal decreases because of dispersion only. These curves are for a signal strength of 186 millivolts per meter at one mile: that is, 1.0 kilowatt of radiated power, equation 5. (Adapted from reference 22.)

parallel to the electric lines of force in the received wave. The ground wave from a vertical antenna is not exactly vertically polarized. The losses in the earth cause the wave front to tilt forward slightly.

Sky-Wave Propagation. Sky-wave reflection occurs at very low frequencies (10 to 30 kilocycles) and into the low-frequency range (30 to 300 kilocycles). The waves at the lower frequencies are propagated through space somewhat as if bounded by two spherical reflecting shells, the ionosphere and the earth.[6] Energy loss in both the earth and the ionosphere is small.

Absorption by the ionosphere becomes important, *particularly in daytime,* in the medium-frequency band (300 to 3000 kilocycles) and the upper part of the low-frequency band. At the standard amplitude-modulation frequencies of 550 to 1600 kilocycles, only the ground wave need be considered in the daytime. During daytime, absorption occurs in the *E* layer. During the night, and particularly in the winter, the decrease in density of ionization and the resulting reduction in absorption by the *E* layer cause reflection to occur. It is difficult to calculate with accuracy the strength of the sky-wave signal, but it can be estimated,[6] assuming a reflection coefficient of about 0.25. Curves are given in reference 16 by which sky-wave field intensities can be determined for the standard broadcast band.

If a radio station is receiving a signal by both the ground-wave path and the sky-wave path, then the received signal is the vector sum of the two components. Sky-wave signals are erratic, and, whenever they are involved, variations in signal strength occur. This is called **fading** and is defined[1] as "variation of radio field intensity caused by changes in the transmission medium." Sometimes, **selective fading** occurs,[24] defined[1] as "fading which is different at different frequencies in a frequency band occupied by a modulated wave."

In the high-frequency band (3 to 30 megacycles) the ground wave is absorbed so rapidly that usually only sky-wave transmission is important. An exception is for such purposes as city police radio systems and installations where the antennas are at a distance above the earth. In such instances space-wave transmission is important. But, for strictly sky-wave propagation, the transmission is so erratic that calculations are unreliable, and curves and maps with field-strength contours are used,[6, 8, 22] an example being Fig. 10.

In fact, as Dellinger has stated,[8] "The more one views the complexities of radio transmission via the ionosphere, the more he marvels that it provides any intelligible communication."* Only because of exhaustive studies[25, 26, 27] over long periods has sky-wave communication been made useful.

Depending on conditions, above about 30 megacycles reflection from the ionosphere becomes uncertain, and above about 60 megacycles reflection from the ionosphere rarely occurs. Thus, at very high frequencies (30–300 megacycles) transmission is normally by the space wave.

Space-Wave Propagation. The space wave was discussed on page 444 and, as indicated in Fig. 3, is composed of a direct wave (or ray)

* Quoted by permission, courtesy J. H. Dellinger, and the American Institute of Electrical Engineers.

and a ground-reflected wave. From the definition on page 444, space-wave transmission is, of course, a special case of ground-wave transmission where the surface wave is attenuated so rapidly that it is neglected.

The signal strength at the receiving antenna is the vector sum of the two waves. The direct wave travels through the atmosphere and is affected by the troposphere. The ground-reflected wave suffers absorption and phase shift at the surface of the earth (page 451).

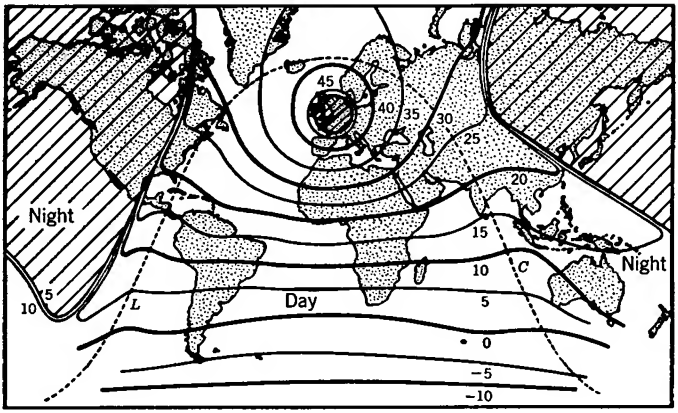


FIG. 10. Lines of equal average received intensity for a particular frequency, time of day, season, epoch of solar cycle, and location of transmitter. The numbers are field intensities expressed in decibels above 1 microvolt per meter, for 1 kilowatt radiated. The dotted line is the locus of sunrise (L) and sunset (C). The shaded areas are regions of essentially zero reception. The transmitting antenna is located in the upper center in the small shaded circle. The frequency is 18.8 megacycles, and the time is noon at the transmitter on a winter day. (From reference 8.)

The length of the path and the heights of the antennas are other factors affecting the received signal strength. The wavelength is short at the frequencies under consideration (30 to 300 megacycles, and 10 to 1 meters), and the location of the antennas should be chosen with care. Once the site is selected, the exact location should be checked experimentally to determine the position of the strongest received signals. Methods of calculating the approximate signal strength with elevated antennas are summarized in reference 6. Reference 28 considers the effect of atmospheric conditions, hills, and buildings.

The following equation[23] for field strength applies approximately

within the "radio path" horizon, equation 10,

$$E = \frac{14\sqrt{w}}{d} \sin\left(\frac{2\pi h_t h_r}{\lambda d}\right) \text{ (volts per meter)}, \tag{9}$$

where w is the power in watts radiated, h_t is the height of the transmitting antenna in meters, h_r is the height above the earth of the receiving antenna in meters, λ is the wavelength in meters, and d is the distance in meters between the antennas. This equation holds for small dipoles (page 460), for distances much greater than the antenna heights and for both vertical and horizontal polarization.

Channels for frequency-modulation broadcast stations start at 88 megacycles and extend to 108 megacycles. Certain television channels also are in this vicinity. For data regarding frequency-modulation signal strengths at various distances, for different radiated power, and for various antenna heights, reference 29 should be consulted.

Direct-Wave Propagation. Only in free space can direct-wave propagation occur but direct-wave propagation can be approximated in regions remote from the earth.

Point-to-point transmission, as from one microwave radio-relay station (or repeater station) to another in a superhigh-frequency radio communication system such as exists from New York to Boston (page 527), is not exactly direct-wave propagation, but it approaches it.

Methods are available[28] for calculating the transmission to be expected at ultrahigh and superhigh frequencies. Many factors must be considered such as frequency, types, and heights of antennas, terrain including trees, hills, buildings, etc.

In the New York–Boston system, operating at about 4000 megacycles and using highly directional antennas located about 30 miles apart on hills, it has been found that ground reflection and variable atmospheric refraction affect transmission.[30] However, it is reported that it is possible during non-fading periods to obtain essentially free-space conditions.

Point-to-point radio transmission at ultrahigh (300 to 3000 megacycles, 100 to 10 centimeters) and superhigh (3000 to 30,000 megacycles, 10 to 1 centimeters) frequencies is possible beyond the **optical,** or **line-of-sight, path,** that is, beyond the horizon. This is due in a large measure to atmospheric refraction and diffraction. The "horizon" for the radio path, or the maximum possible "direct-ray" path is

$$d = \sqrt{2h_t} + \sqrt{2h_r}, \tag{10}$$

where d is the distance in miles, and h_t and h_r are the heights in feet

above the surface of the earth of the transmitting and receiving antennas.

Atmospheric "ducts" as they are called are caused by refraction in the troposphere and reflect superhigh-frequency radio waves back to the earth, where they are reflected back to the refracting region, and so on. Thus, radio waves are "trapped" within the duct, and peculiar transmission phenomena occur.[31] Transmission at these frequencies is also affected quite adversely by rain, particularly at wavelengths below 5 centimeters.[32, 33]

Antennas, or Aerials. The pages that follow will deal largely with transmitting antennas, and little need be said about receiving antennas. This is because of the application, by Carson, of the reciprocity theorem (page 151) to antennas. This theorem, often called the **Rayleigh-Carson reciprocity theorem,** because it was due to Rayleigh in its original form, may be stated: *If a radio-frequency voltage E, when inserted at a given point x in antenna 1, causes a current I to flow at a given point y in distant antenna 2, then the same voltage E when inserted at point y in antenna 2 will cause the same current I to flow at point x in antenna 1.* This theorem holds only if the transmission path remains constant. If a given antenna has certain directional properties in transmitting, according to this theorem it has the same directional properties in receiving, and vice versa.

There is another important reason for concentrating on transmitting antennas. They are usually designed and constructed to radiate large amounts of power efficiently and to operate at a single narrow frequency band. On the other hand receiving antennas often must operate over a rather wide band and should not be efficient at one part of the band at the expense of other parts.

An **antenna, or aerial,** is defined[34] as "a means for radiating or receiving radio waves." Many types, shapes, and lengths of transmitting and receiving antennas are used, and they have various positions with regard to the surface of the earth and objects such as buildings. Attention will be focused on basic types, first as they behave in free space, and then as their characteristics are modified by the presence of the earth.

In studying antennas, it is convenient to use the classification periodic antennas and aperiodic antennas, the latter being discussed on page 479.

Periodic Resonant Antennas. Periodic antennas are of many types and constitute a large percentage of all antennas used. **Periodic antennas** are defined[34] as "those in which the impedance varies as the frequency is altered due to reflections or standing waves within

the antenna system. This includes open-end wires and resonant antennas of all kinds."

A simple antenna was shown in Fig. 1. From the electric-circuit standpoint such an antenna is far more than two wires, as Fig. 11 indicates. If the antenna is driven with a voltage of frequency such that the *total antenna length* is approximately one-half wavelength, the current (I) and voltage (E) distribution along the antenna will be as

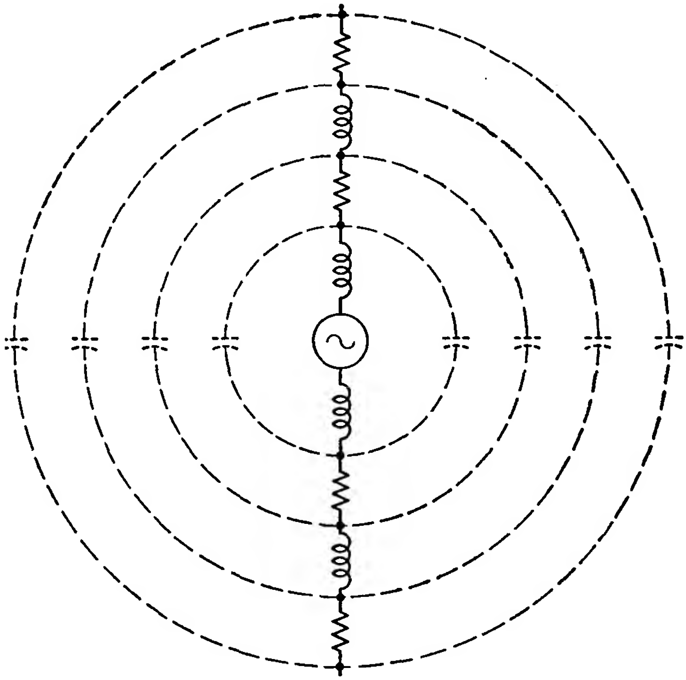


FIG. 11. An antenna is approximately equivalent electrically to the circuit shown.

shown in Fig. 12(a). This is similar to the stationary (standing) wave on a transmission line (Chapter 6) but differs in one important respect. When a theoretical *lossless* line is considered, the current and voltage distribution drop to zero at certain points. But, even if an antenna were made of lossless wire, when it was driven, it would radiate energy into space. Therefore, the current and voltage distribution would not be zero at certain points as for a lossless line, except that the current would be zero at the extreme end of each wire. This means that the input impedance of an antenna must always have a finite resistance component.

Confusion exists as to what is represented by the curves of Fig. 12(*a*). The dotted current curve *I* represents the *effective* values of current that would be indicated by radio-frequency ammeters connected at various points *in* the antenna wires. These *effective* current values (and also voltage values) are plotted with magnitude of current along the horizontal or x axis, marked "center line" in Fig. 12(*a*). The solid voltage curve *E* represents the *effective* values of voltage that would be indicated by (imaginary) voltmeters connected between the various points on the antenna and a conducting plane passing through the center of the antenna and at right angles to it (along the center line).

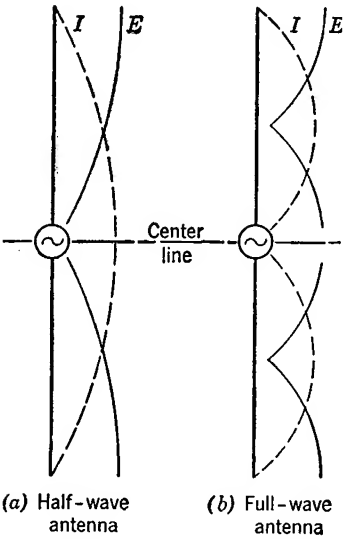


FIG. 12. The current *I* and voltage *E* will be distributed on the antennas approximately as shown. Because the antennas radiate power, the curves do not reach zero (except at the ends), and the input impedance always has a finite resistance component.

From the discussion given in Chapter 6, it follows that the input impedance of the antenna of Fig. 12(*a*) is a low value of pure resistance (low voltage, high current) and that the input impedance of Fig. 12(*b*) is a high value of pure resistance (high voltage, low current). If the driving frequency is such that the antenna is neither a half wave nor a full wave in length, then the input impedance will be composed of both resistance and reactance. The velocity of propagation along an antenna is not exactly the propagation in free space. The actual antenna length should be *about* 0.95 times the theoretical length. Also, the distribution of current and voltage is not exactly sinusoidal. Because the input impedance is a low value of pure resistance (like a series resonant circuit) at certain frequencies and a high value of pure resistance (like a parallel resonant circuit) at other frequencies, these antennas are often called **resonant antennas.**

Free-Space Radiation Patterns from Periodic Straight-Wire Antennas. One type of antenna used extensively is called a **doublet antenna** defined[35] as "an antenna consisting of two elevated conductors substantially in the same straight line of substantially equal length, with power delivered at the center." Such an antenna is often called a **dipole.**[1]

In deriving the radiation pattern for a periodic straight-wire doublet or dipole antenna, the equation for the radiation from an **elementary doublet,**[3] of such small length that the current in all parts is the same, is used. The next step in determining the radiation pattern of an actual antenna is to assume the actual antenna to be made up of a number of elementary doublets and to determine graphically or mathematically[3] the field intensity produced by all the elementary doublets at the various points in space about the antenna under study.

For periodic straight-wire doublets, or dipoles, the *free space* radiation patterns are[36]

For a periodic straight-wire antenna an odd number of half wavelengths long:

$$E = \frac{\dfrac{60I}{d}\cos\left(\dfrac{\pi l}{\lambda}\cos\theta\right)}{\sin\theta}. \tag{11}$$

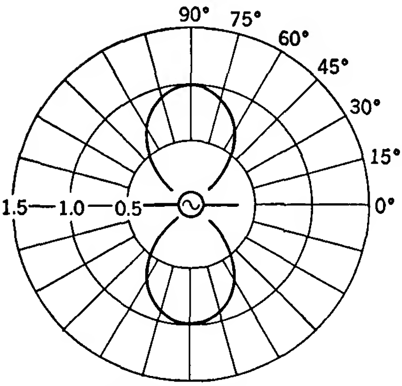
For a periodic straight-wire antenna an even number of half wavelengths long:

$$E = \frac{\dfrac{60I}{d}\sin\left(\dfrac{\pi l}{\lambda}\cos\theta\right)}{\sin\theta}. \tag{12}$$

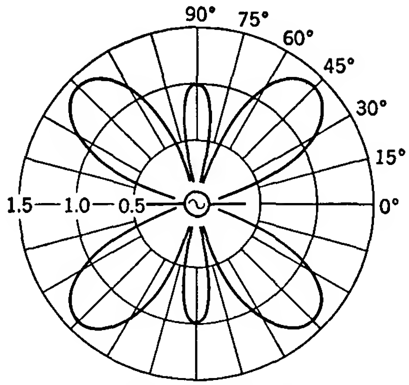
In these equations E is the effective value (r.m.s.) of the field strength in volts per meter at the point under consideration, d is the distance in meters from the point to the antenna, I is the effective value of the current at the point of maximum current flow in the antenna, l is the antenna length in meters, λ is the wavelength in meters, and θ is an angle measured as shown in Fig. 13.

These figures show the shapes of the *free-space* radiation patterns for several straight-wire antennas. The relative field strength of the signal radiated at any angle is the length of a vector at that angle from the center of the antenna to the pattern edge. Information is available[36] for finding the positions and relative magnitudes of the lobes without first plotting the figures. In Fig. 13 is shown the radiation pattern in a plane parallel to the antenna, and in which the antenna lies. The entire radiation is a "solid figure," obtained by "revolving" the patterns shown about the antenna as an axis.

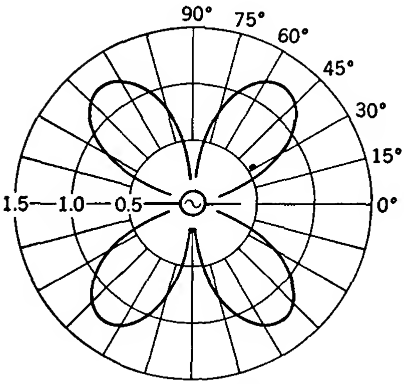
Ground-Reflection Factors. The radiation patterns of Fig. 13 are for *free space* and apply in practice only if *all* reflections are neglected. When antennas are near the earth, as they often are, the effect of reflections from the earth must be considered. To obtain an approximate idea of the actual radiation pattern, several assumptions are made: *first,* that vertically polarized waves are reflected from the earth *without* a change in phase; *second,* that horizontally polarized waves are reflected from the earth *with* a change in phase of 180°; and, *third,* that no energy is lost during reflection. If these assump-



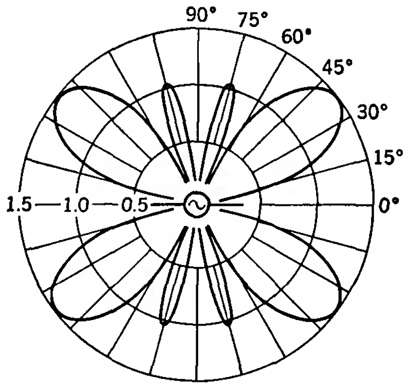
(*a*) Antenna length λ/2



(*c*) Antenna length 3/2 λ



(*b*) Antenna length λ



(*d*) Antenna length 2 λ

Fig. 13. Free-space radiation patterns plotted from equations 11 and 12. Relative magnitude of radiated field strength is indicated by 0.5, 1.0, and 1.5. Areas of figures should *not* be compared. These figures represent radiation in one plane. The complete radiation is a "solid figure" obtained by revolving the pattern about the antenna (at the center) as an axis.

tions are made, it is then possible to determine theoretical **ground-reflection factors,** which when applied to the free-space radiation pattern give the (theoretical) antenna radiation pattern in the presence of the earth. Note in particular that in this section *plots of ground-reflection factors and not antenna radiation patterns are being developed.*

Because the reflected wave is assumed to be shifted in phase either 0° or 180° at the surface of the earth, it is possible to represent a reflected wave as coming from an **image antenna** as in Fig. 14. Half-

wave antennas are employed as illustrations because they are relatively simple. The method can be extended to longer antennas and to antennas that are other than vertical or horizontal.[6, 21]

In Fig. 15 are shown plots of ground-reflection factors for various conditions. As an illustration of how these are determined, typical calculations will be made for a horizontal half-wave antenna $\frac{1}{4}\lambda$ (or 90°) above the surface of the earth and at an angle of 30° with the horizontal. The relative radiation at some distant point, where the rays

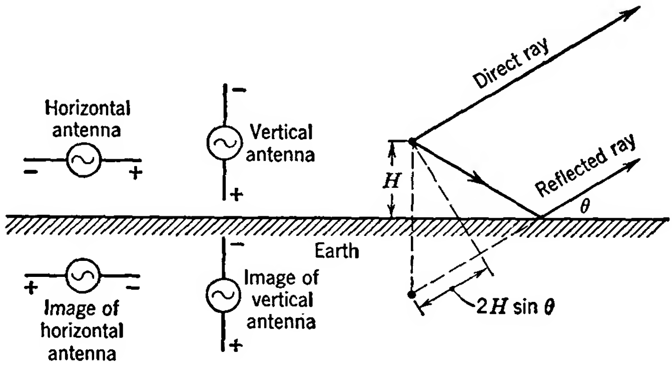
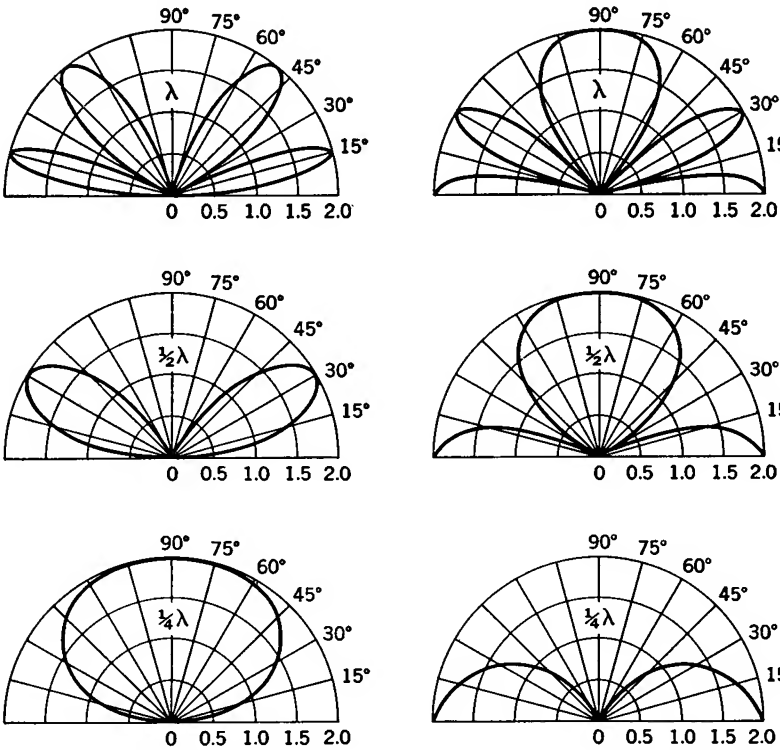


Fig. 14. Because a horizontally polarized wave is *assumed* to be reflected with a 180° shift in phase and a vertically polarized wave is assumed to be reflected with 0° shift in phase, the action of the reflected wave for horizontal and vertical antennas can be represented as if an image of the antenna, or an "image antenna," were an equal distance below the surface of the earth. The polarity of an image antenna is the same as would be induced in a wire at that point. To reach a distant point, the reflected ray, assumed to come from the image antenna, must travel a distance $2H \sin \theta$ farther than the direct ray. Half-wave antennas are shown because they are common.

are assumed to arrive parallel, will be computed. The ground-reflected ray travels (Fig. 14) $2H \sin \theta = 2 \times 90° \sin 30° = 90°$ farther than the direct ray; also, there is an *assumed* 180° shift in phase at the instant of reflection. (An alternate viewpoint is that the signal from the image antenna of Fig. 14 is 180° out of phase.) Thus, at the distant point, at 30° with the horizontal, the ground-reflected ray is $90° + 180° = 270°$ behind the direct ray which is equivalent to leading by 90°. Assuming no loss at reflection and that the magnitude of each ray is 1.0, the magnitude of the signal is the vector sum of these two rays or 1.414. This is plotted at 30° in the lowest diagram of Fig. 15(a). Other factors are determined in a similar manner. For the lowest diagram of Fig. 15(b), at 30° the ground-reflected ray will



(a) Horizontal antennas *(b)* Vertical antennas

FIG. 15. Ground-reflection factors for the vertical and horizontal half-wave antennas of Fig. 14. The wavelength notations on each figure are the height of the antenna above the surface of the earth. These are *not* radiation patterns.

travel $2H \sin \theta = 90°$ farther, and, since there is an *assumed* shift of 0° at the surface, the reflected ray will arrive at the distant point 90° behind the direct ray. Assuming no loss at reflection, the resultant field will be the sum of two vectors of magnitude 1.0 and 90° out of phase, and the reflection factor will be 1.414 as before.

Equations can be written which will give the reflection factors. For the diagrams of Fig. 15(*a*) the equation is

$$\text{Reflection factor} = 2 \sin (H \sin \theta), \tag{13}$$

which applies to horizontal half-wave antennas of any height H in degrees above the earth. For the diagrams of Fig. 15(*b*) the equation is

$$\text{Reflection factor} = 2 \cos (H \sin \theta), \tag{14}$$

which applies to vertical half-wave antennas with centers at any height

H above the surface of the earth. Plots of reflection factors for many heights are available.[37]

Radiation Patterns from Periodic Straight-Wire Antennas near the Surface of the Earth. If the free-space radiation pattern (Fig. 13) of an antenna is modified by the application of ground-reflection

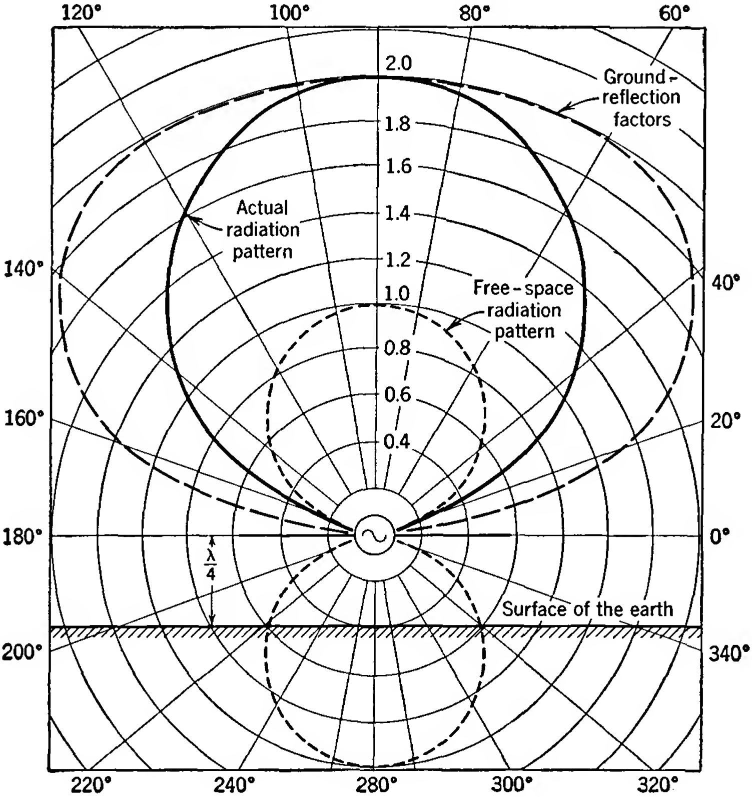


FIG. 16. To obtain the radiation pattern in the vertical plane (in which the antenna lies) for a horizontal half-wave antenna $\frac{1}{4}\lambda$ above the surface of the earth, multiply the free-space radiation (Fig. 13) at various angles by the ground-reflection factor (Fig. 15) at that angle. The areas of these figures should be disregarded.

factors (Fig. 15), or by equation 13 or 14, the theoretical radiation pattern of an antenna near the earth is obtained. In the following discussion half-wave antennas will again be considered, and the radiation in the most important planes will be depicted.

Horizontal Half-Wave Antennas, Vertical-Plane Radiation. The plane to be considered in this paragraph will be the *vertical plane in*

which the antenna lies. As an example of the method of calculation, Fig. 16 has been included, which shows how the solutions are made for a *half-wave antenna* $\frac{1}{4}\lambda$ *above the earth.* Thus, at 60°, the free-space radiation is 0.8, the reflection factor is about 2.0, and the theoretical

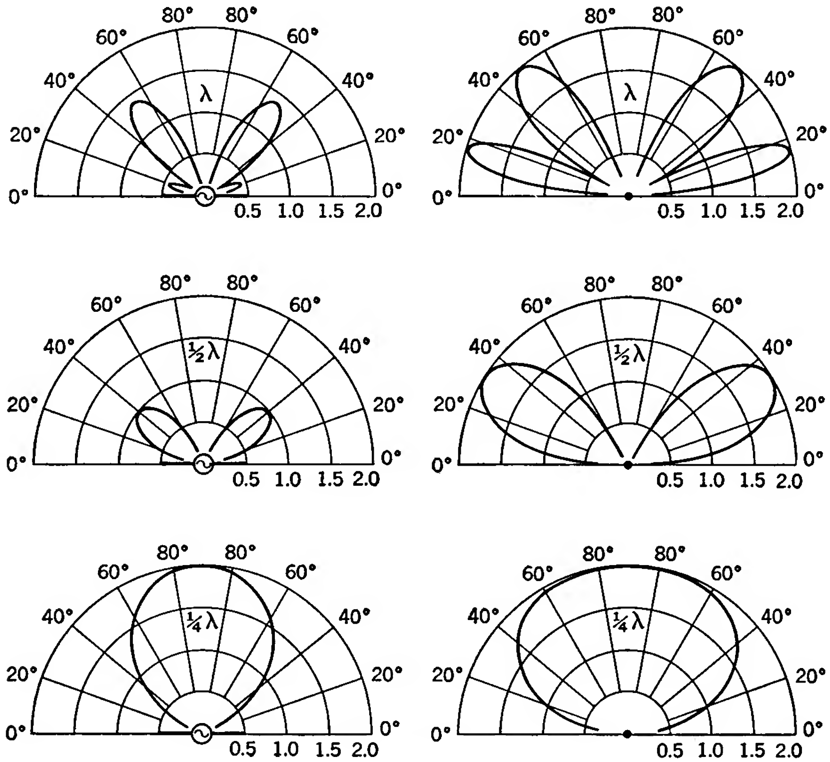


FIG. 17. Radiation patterns for *horizontal* half-wave antennas at $\frac{1}{4}\lambda$, $\frac{1}{2}\lambda$, and λ above the earth. Figures at the left are for radiation in a vertical plane in which the antenna lies, and those at the right are for a vertical plane passing through the center of the antenna at right angles to the antenna.

radiation at 60° is $0.8 \times 2.0 = 1.6$, as shown in Fig. 16. The radiation at other angles is found in the same way. This same method is used to determine the radiation patterns (Fig. 17) for half-wave antennas at heights other than $\frac{1}{4}\lambda$.

The radiation pattern to be considered in this paragraph is for a *plane passing at right angles through the center* of a horizontal half-wave straight-wire antenna. The pattern will have the same shape as the plot of the ground-reflection factors of Fig. 15(*a*) for the antenna

height under consideration (Fig. 17). This is because in free space the radiation as "viewed" from the end of a half-wave straight-wire antenna is a circle.

Horizontal Half-Wave Antennas, Horizontal-Plane Radiation. Theoretically, this antenna radiates no energy in a horizontal plane in which the antenna lies, because the reflection factors at a vertical angle of 0° are zero, as Fig. 15(*a*) indicates. But, at many vertical angles there is radiation, and a plot can be made[37] of field intensities at these vertical angles in the various horizontal, or azimuth, directions from the antenna. Referring to Fig. 16, at 40° the radiation in the vertical plane in which the antenna lies is approximately 1.1 as determined from the "actual radiation pattern." At 40° at right angles to the

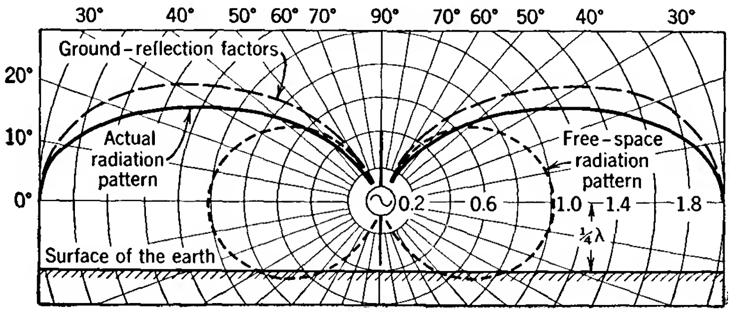


FIG. 18. To obtain the radiation pattern in the vertical plane (in which the antenna lies) for a vertical half-wave antenna with its center $\frac{1}{4}\lambda$ above the surface of the earth, multiply the free space radiation (Fig. 13) at the various angles by the ground-reflection factor (Fig. 15) at that angle. The areas of these figures should be disregarded.

center of the antenna the radiation is approximately 1.7 as determined from the "ground reflection factors," which, as mentioned in the preceding paragraph, is also the radiation pattern in a plane passing through the center of the antenna at right angles.

Vertical Half-Wave Antennas, Vertical-Plane Radiation. The plane to be considered is any *vertical plane in which the antenna lies.* As an example, Fig. 18 has been included, which shows how the calculations are made for a *half-wave antenna the center of which is $\frac{1}{4}\lambda$ above the earth.* Thus at 30° the free-space radiation is about 0.8, the reflection factor is about 1.4, and the theoretical radiation at 30° is $0.8 \times 1.4 = 1.12$, as shown in Fig. 18. The radiation at other angles is found in the same manner. This same procedure is used to determine the radiation patterns for half-wave antennas at heights other than $\frac{1}{4}\lambda$.

Vertical Half-Wave Antennas, Horizontal-Plane Radiation. The radiation pattern in a *horizontal plane* for a vertical half-wave straight-wire antenna at any height above the earth will be a circle. This can be visualized by "looking down" on Fig. 18. The radiation in all horizontal directions is uniform. However, the radiation at a vertical angle of 0° (the surface of the earth) is greater than the radiation at some vertical angle. It should be mentioned that losses in the surface of the earth and in objects on the surface of the earth reduce the radiation along the earth below the values indicated by Fig. 19.

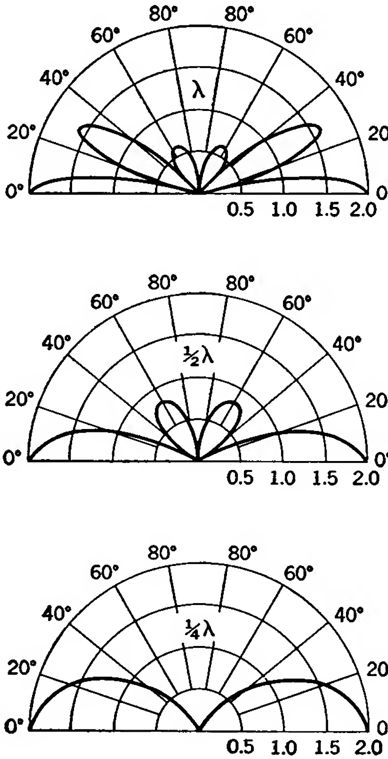


FIG. 19. Radiation patterns for *vertical* half-wave antennas with centers $\frac{1}{4}\lambda$, $\frac{1}{2}\lambda$, and λ above the earth. The antennas are located at the lower centers of the figures and are not shown for simplification.

Directional Periodic Antennas. In many radio applications it is desired that the antenna radiating (or receiving) systems be directional. A **directional antenna** is defined[34] as "an antenna having the property of radiating or receiving radio waves more effectively in some directions than others." Directional antennas are particularly useful with important point-to-point installations because they reduce interference, increase secrecy, and save power, resulting in lower overall first costs and operating charges. Directional transmitting antennas are sometimes called **directive antennas.**

Antenna Gain. The **gain of an antenna** is defined[34] as follows: "The measured gain of one transmitting or receiving antenna over another is the ratio of the signal power one produces at the receiver input terminals to that produced by the other, the transmitting power level remaining fixed." The reference antenna for measuring the gain of a directional antenna is often a similar non-directional radiator. The gain as defined is a power gain and is often expressed in decibels.

Antenna Arrays. Antenna arrays are systems of antennas coupled

together for the purpose of obtaining directional effects.[34] When the unit antennas are arranged in a line perpendicular to the direction of transmission the system is called a **broadside array,** and when the unit antennas are arranged in the direction of transmission it is called an **end-on array, end-fire array,** or an **alignment array.**[34]

Broadside Arrays. One broadside array consists of half-wave antennas arranged along the same straight line and is sometimes called the **collinear array.** (See also reference 34, linear array.) A simple

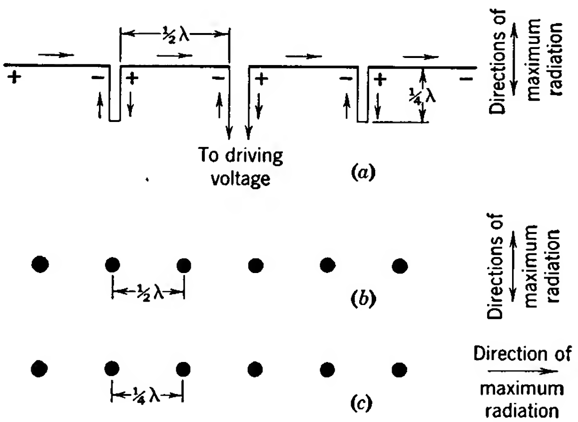


FIG. 20. In (*a*) is shown a collinear array. In (*b*) is shown a broadside array of conductors, the ends of the wires, or radiators, being shown by dots. For the spacing indicated, the conductors are driven in phase. In (*c*) is shown an end-on, or end-fire, array. For the spacing shown each radiator is driven −90° out of phase with respect to the radiator on the left. This array has maximum radiation in one direction only.

arrangement is shown in Fig. 20(*a*). When the driving voltage has the polarity shown, the instantaneous current directions will be as indicated by the arrows. As a result, cancellation occurs for the quarter-wave sections, and the half-wave antennas combine to send signal at right angles to the direction in which they point, giving a free-space radiation pattern similar to Fig. 13(*a*), but very much more elongated. An array such as Fig. 20(*a*) could be horizontal or vertical (or at any other angle) with respect to the earth. Reflection from the earth would affect the radiation pattern. Radiation from the antenna of Fig. 20(*a*) will occur in both directions at right angles to the antenna, hence the word "broadside."

Another broadside array (usually called the **broadside array**) is shown in Fig. 20(*b*) in which a number of half-wave antennas are

arranged with their axes at right angles to the direction of the array. The dots represent the upper ends of the half-wave antennas. They are $\frac{1}{2}\lambda$ (or 180°) apart and are driven in phase. The radiation in the plane of the page is zero for the directions in which the radiators are aligned. A signal from one radiator will arrive at a second radiator 180° after it is radiated. By this time, the applied voltage at the second has changed 180°, and the radiation from the second radiator will cancel that which has arrived from the first. Thus cancellation of signal occurs, and there is negligible radiation in the direction of the array. At points in front and in back of the antenna, the radiations add, giving two directions of maximum radiation as shown in Fig. 20(*b*). This array may be placed at any position with respect to the earth. Its radiation pattern will be affected by reflection from the earth.

End-fire, End-on, or Alignment Arrays. This is usually called[34] an **end-fire array** in practice, a typical arrangement being shown in Fig. 20(*c*). The currents in these radiators (often half-wave dipoles) are equal in magnitude, but the driving arrangement is such that there is a progressive phase lag between adjacent antennas in the direction of maximum radiation. If the radiators are $\frac{1}{4}\lambda$ apart, the phase lag is 90°, etc. However, the theory does not apply if the spacing exceeds $3\lambda/8$. With this arrangement, maximum radiation occurs in the direction of the radiator having the most lagging phase.

Reflectors and Directors. It will be noted that the broadside arrays of Fig. 20(*a*) and (*b*) radiate equally in both directions, usually a very objectionable feature. They are made to radiate essentially in one direction only by the use of reflectors, directors, or both.

An **exciter**[34] is "the portion of a transmitting array, of the type which includes a reflector, which is directly connected with the source of power."

A **reflector**[34] is "a parasitic element located in a direction other than the general direction of the major lobe of radiation."

A **director**[34] is "a parasitic element located in the general direction of the major lobe of radiation."

An antenna array with a reflector is shown in Fig. 21. If a second identical row of conductors is placed $\frac{1}{4}\lambda$, or 90°, from the *driven* row of Fig. 20(*b*), but *not* connected to it or to the source of radio-frequency energy, then this second row will be **parasitically excited** by the driven row. A signal leaves a driven exciter of Fig. 21 and arrives at the corresponding reflector 90° after it is radiated. This signal induces a voltage in the reflector which will cause the radiation of a field of 180° out of phase with the exciting field. This field will

travel in two directions, one of which is toward the exciter where it arrives 90° after being radiated by the reflector. During this time, the signal voltage driving the exciter has changed 180°, and, hence, the new signal coming from the exciter and the signal from the reflector are in phase and *add* in the east direction. In the west direction, the original signal and the parasitically radiated signal are 180° out of phase, and radiation largely is *cancelled* in this direction. Thus radiation occurs largely in the east direction as Fig. 21 indicates. Antennas of this type were formerly used extensively in trans-oceanic radio-telephone systems, but they have been superseded by rhombic antennas (page 480). A simple arrangement for feeding the driven row of Fig. 21 is shown in Fig. 22.

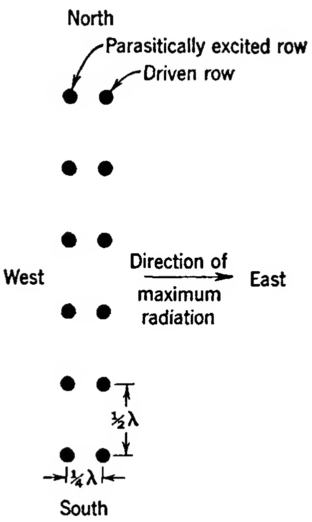


FIG. 21. A directional antenna array consisting of a front row that is driven in phase, and a back row that is excited parasitically. The dots represent the ends of the conductors.

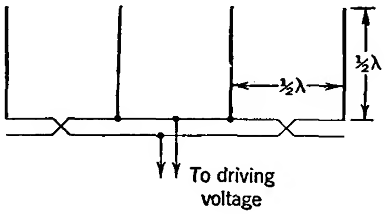


FIG. 22. Side view of a portion of the driven row of the antenna array of Fig. 21, showing how the radiators are connected to obtain the correct phase relations.

One or more directors (and a reflector as well) are used with the **Yagi antenna,** named after its inventor. The arrangement of a typical array is shown in Fig. 23; it is often mounted horizontally and arranged for rotation. Spacings of less than 90° are used, and the individual conductors are not all the same length. Some design data are available, but spacings and lengths are usually determined experimentally.[37, 39] A reflector is usually made longer than the exciter, and the directors are usually made shorter. Only one reflector is usually employed, but sometimes several directors are used.

Antennas for Amplitude-Modulation Broadcast Stations. These antennas operate at fixed frequencies within the band 550 to 1600

kilocycles. In general, they are designed to supply a strong signal to the area adjacent to the transmitter, transmission being largely by the surface-wave component of the ground wave. Usually, the coverage of the area is merely attributed to the ground wave.[16] The antenna radiates the carrier and two sidebands (page 489), requiring about 5000 cycles each side of the carrier. The carrier frequency is used in calculations.

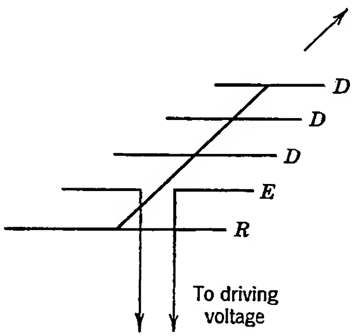


FIG. 23. A Yagi directional array using three directors (D), one reflector (R), and one exciter (E). This antenna is often mounted on top of a pole and arranged for rotation. The line at the center of the antenna elements is the support.

Non-directional Antennas for A–M Broadcast Station. Because ground-wave propagation is desired, the antennas usually are vertical and radiate vertically polarized waves. In general, it is desired that sky-wave radiation be a minimum. In daytime sky-wave transmission at 550 to 1600 kilocycles is negligible because of ionospheric absorption (page 445), but, at night, reflection from the ionosphere occurs, and one broadcast station may seriously interfere with another station, perhaps a thousand miles away and operating on the same frequency or an adjacent frequency. In the design of a broadcast antenna, it is, therefore, necessary to consider the radiation both in the horizontal plane, represented by the surface of the earth, and in vertical planes.

If the antenna system is to be non-directional, a single vertical antenna is used. Neglecting losses, the radiation along the surface of the earth will be uniform (page 444), giving a circular pattern. The method of computing the field intensity is discussed on page 453. The shape of the radiation pattern in a vertical plane depends on the antenna height, the patterns for several heights being shown in Fig. 24. Although an antenna height of about 0.62λ theoretically gives the strongest ground-wave signal,[40] it will be noted that a lobe has developed at about 60° with the horizontal. Such a lobe causes undesired high-angle sky-wave radiation that wastes power and may cause distant interference. When all factors are considered, a straight vertical antenna about 0.53λ has been found to make an excellent radiator.[41] At the lower end of the broadcast band, an antenna this high is, physically, a sizable and expensive structure, and, although such towers are used, they are often a smaller fraction of a wavelength.

Many investigations have been made regarding height and shape of

broadcast antennas, and the ground systems to be used with them. For many broadcast purposes an antenna of height about 0.25λ or even less, with a good ground system, has been shown[12] to be satisfactory. A good ground system for a vertical antenna often consists of about 120 heavy copper wires, each about 0.25λ to 0.5λ long, and buried at a depth of several inches so that they extend radially from the base of the antenna.

Directional Antennas for A–M Broadcast Stations. Sometimes it is necessary to use a directional antenna system with a broadcast station. For instance, sometimes the geographical distribution of the population

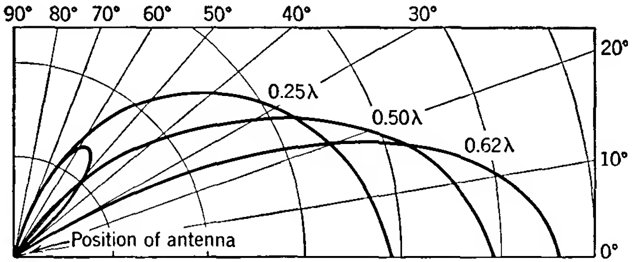


FIG. 24. Radiation patterns in the vertical plane passing through vertical antennas of height 0.25λ, 0.50λ, and 0.62λ. The lobe at about 60° is for the 0.62λ antenna. The diagrams are one half of the pattern which is symmetrical about the antennas.

is such that this is desirable; or perhaps it is necessary to "protect" the service area of a distant station by designing the radiation pattern so that the signal sent toward that station is not too strong. Such directional antenna systems usually consist of two or more vertical antennas located a fraction of a wavelength apart and driven by the same transmitter. For explaining the basic principle of operation, a two-element array will be considered.

The design of a directional broadcast antenna system which will produce a given radiation pattern is a cut-and-try procedure. Thus, suppose that, after a careful study is made of all factors, an antenna radiation pattern such as Fig. 25 is selected. References 6 and 43 to 46 present many patterns and the means of producing them. Assume that an investigation indicates that to obtain a pattern such as Fig. 25 the two towers should be spaced 135° apart (one wavelength equals 360°), that the current in antenna B should lag that in A by 50°, and that the current in antenna B should be 0.7 that in A. These data will now be checked.

From Fig. 26 it is evident that radiation from antenna B must travel a distance $S \cos \theta$ farther than the radiation from antenna A to

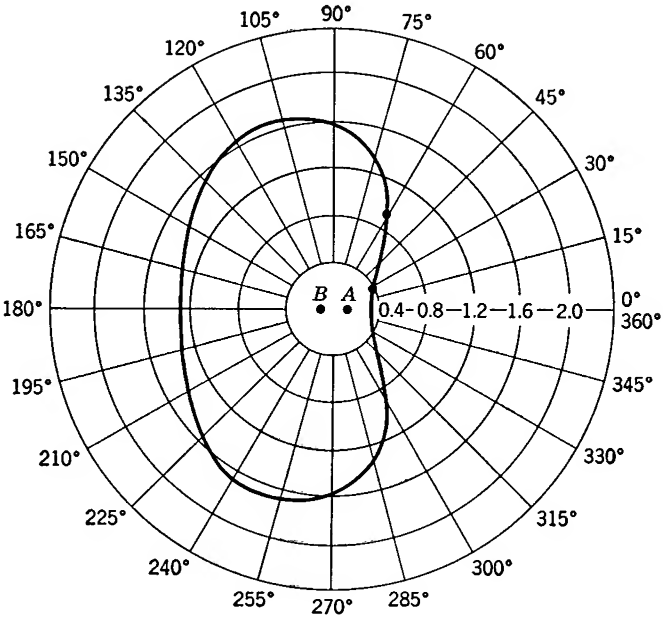


FIG. 25. Radiation pattern in a horizontal plane at the surface of the earth for two vertical antennas A and B that are 135° apart and driven so that the current in B is 0.7 of that in A and lags by 50°.

reach some distant point. Using the angle $\theta = 30°$ as an illustration, $S \cos 30° = 135° \times 0.866 = 117°$. But the signal from antenna B is radiated 50° behind that from A; hence at the distant point under consideration the signal from antenna B arrives $117° + 50° = 167°$ behind the signal from A. Thus, the total signal strength relative to A alone at an angle of 30° is the sum of two vectors, the first is assumed to be unity in magnitude and to be the reference vector, and the second is assumed to be 0.7 in magnitude ($I_B = 0.7\, I_A$) and lags 167°. The signal strength at the distant point relative to that produced by antenna A alone is

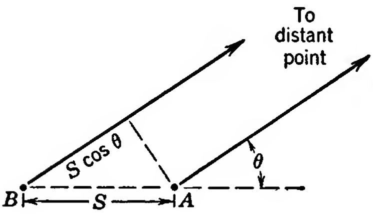


FIG. 26. The signal from vertical antenna B must travel a distance $S \cos \theta$ farther than the signal from A to reach a distant point in the horizontal plane.

$$\sqrt{(1.0 + 0.7 \cos 167°)^2 + (0.7 \times \sin 167°)^2} = 0.35.$$

This value is then plotted as indicated in Fig. 25, and other values are calculated to give the complete radiation pattern.

An equation for the relative radiation compared to that due to antenna A in a horizontal plane at the surface of the earth at various bearing, or azimuth, angles θ, measured with respect to a line joining the antennas as shown in Fig. 26, readily follows from the preceding reasoning. This equation is

Relative field strength at angle θ =

$$\sqrt{[1 + k\cos(S\cos\theta + \alpha)]^2 + [k\sin(S\cos\theta + \alpha)]^2}, \qquad (15)$$

where k is the ratio of current in antenna B to that in A, α is the angle by which the current in antenna B lags that in A, S is the spacing in degrees, and θ is as previously noted. Thus, at 60°, the array previously considered produces a relative field strength of

$$\sqrt{[1 + 0.7\cos(135^\circ\cos 60^\circ + 50^\circ)]^2 + [0.7\sin(135^\circ\cos 60^\circ + 50^\circ]^2} = 0.92.$$

This is found to fall on the curve of Fig. 25.

The calculations just given are for the relative radiation only, compared to antenna A that is assumed to give unit radiation. If the actual field strength at any angle is desired, it can be computed for antenna A alone from equations 6, 7, or 8 and as discussed in the paragraph just following these equations. The radiation for antenna *A alone,* multiplied by the factors at the various angles from Fig. 25, will give a radiation pattern for the antennas in millivolts per meter, or other units used.

Thus far, only the radiation, due to the two antennas, in a horizontal plane at the surface of the earth has been considered. Often, it is desired to know the radiation at various angles and in various vertical planes, so that the signal strength at some distant point via the sky-wave path can be determined.[16] For this purpose, the following equation is often used[41, 44]

Relative field strength =

$$\sqrt{1 + k^2 + 2k\cos(\alpha + S\cos\theta\cos\phi)}\left[\frac{\cos H - \cos(H\sin\phi)}{(\cos H - 1)\cos\phi}\right]. \qquad (16)$$

In this equation ϕ is the angle in the vertical plane at which the radiation is to be computed and is measured with respect to the horizontal, H is the height of each of the two identical antennas in degrees, and all other values are as previously explained. Thus, suppose the height of each of the antennas previously considered is 65°,

or less than one-fourth wavelength (90°), and the radiation along the horizontal ($\theta = 0°$) at an angle of 60° with respect to the line of the towers is desired. Then, from equation 16,

$$\sqrt{1 + (0.7)^2 + 2 \times 0.7[\cos(50° + 135 \cos 60° \cos 0°)]} = 0.92.$$

At a zenith angle of 30° with respect to the horizon, and at an azimuth angle of 30° with respect to a line joining the two towers, the relative radiation will be from equation 16,

$$\sqrt{1 + (0.7)^2 + 2 \times 0.7[\cos(50° + 135° \cos 60° \cos 30°)]} \times \left[\frac{\cos 65° - \cos(65° \sin 30°)}{(\cos 65° - 1)\cos 30°}\right] = 0.86$$

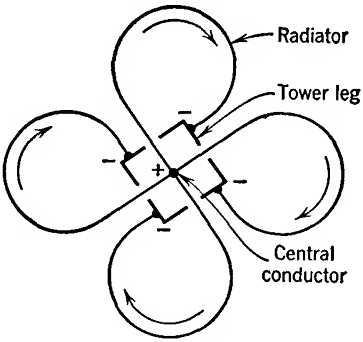


FIG. 27. One radiating unit consisting of four loops, used in a frequency-modulation broadcast antenna array. The array consists of, perhaps, six or eight (as desired) elements as shown placed in a horizontal position on a vertical tower, at intervals of $\frac{1}{2}\lambda$. An end view of the tower is shown. The tower legs and a conductor at the center feed the loops as indicated, the arrows being instantaneous current directions. Sometimes three legs and three loops are used.

Antenna patterns can be determined by mechanical and electrical devices.[47, 48, 49]

The preceding discussion was for directional antenna arrays of two vertical towers. Sometimes, more than two towers are used. Information on such arrays can be found in the references already listed, particularly in reference 46.

Directional Antennas for F–M Broadcast Stations. From the standpoint of size and shape, antennas used in frequency-modulation broadcasting differ greatly from those used in amplitude modulation. Among the reasons are *first,* the frequency is about 100 times greater, making possible antennas with small dimensions; *second,* horizontal polarization is used with frequency modulation instead of vertical polarization as in amplitude modulation; *third,* because there is less tendency for frequency-modulation stations to interfere and because transmission by sky wave does not usually occur at the high frequencies assigned, antenna systems that are horizontally directional are employed only to a limited extent.

The antennas used in amplitude-modulation broadcast systems are

almost always vertical towers, one being but little different from another. In frequency-modulation on the other hand, many shapes are used for the radiators.

The configuration of the radiators of a typical antenna used in frequency-modulation broadcast is shown in Fig. 27. The loops are conductors of large diameter and are connected as indicated, the tower lags in parallel feeding one end of the loops, and a conductor at the center feeding the other end. Each loop conducts current as indicated by the arrows and is a radiator. As many as eight elements, such as shown in Fig. 27, are sometimes "stacked" into an array.[50]

Driving-Point Impedance of Periodic Antennas.[51] As shown in Fig. 11 an antenna is an electric circuit composed of resistance, inductance, and capacitance, and the voltage applied across its input terminals divided by the current that flows into these terminals is a measure of the **driving-point impedance, self-impedance,** or **input impedance,** as it is often called. As explained on page 460, for periodic antennas this impedance varies with the frequency of the applied voltage.

The reason that the input impedance varies is evident from Fig. 12. From the theory of standing (stationary) waves, the input currents that voltages of different frequencies will produce will vary in both magnitude and phase, and hence the input, or driving-point, impedance will vary.

It is possible to predict the input impedance of antennas of various types from theoretical consideration,[3, 44] but in practice, if the impedance must be known, bridge measurements are usually made on the actual antenna. As will be seen in the chapter that follows, the reactive component of the driving-point impedance is either ignored or canceled out with opposite reactance at the point where the antenna is driven.

Of interest is the **antenna resistance,** defined[34] as "the quotient of the power supplied to the entire antenna circuit by the square of the effective antenna current referred to a specified point." Antenna resistance includes[34] radiation resistance, ground resistance, radio frequency resistance of conductors in the antenna circuit, equivalent resistance due to corona, eddy currents, insulator leakage, and dielectric power loss.

In antenna design, of much importance is the **radiation resistance,** defined[34] as "the quotient of the power radiated by an antenna by the square of the effective antenna current referred to a specified point."

The radiation efficiency can be determined if the antenna resistance

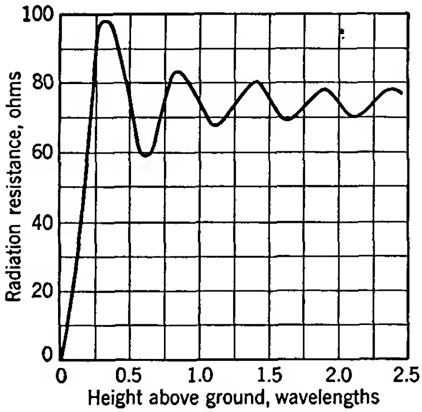


FIG. 28. The radiation resistance of a half-wave antenna in free space is about 73 ohms. When near the earth, the radiation resistance varies as indicated because of reflection.

and radiation resistance are known. **Radiation efficiency** is defined[34] as "the ratio of the power radiated to the total power supplied to the antenna at a given frequency."

The theoretical radiation resistance of a half-wave antenna in free space and driven at center is 73 ohms. The radiation resistance of the horizontal half-wave antenna is affected by its position above the earth as indicated in Fig. 28.

The input impedance of a typical vertical antenna for amplitude-modulation broadcast service is shown in Fig. 29. This antenna has zero reactance and hence is said to be **resonant** at frequencies of 0.685, 1.070, and 1.715 megacycles. At 0.685 megacycles the resistance is about

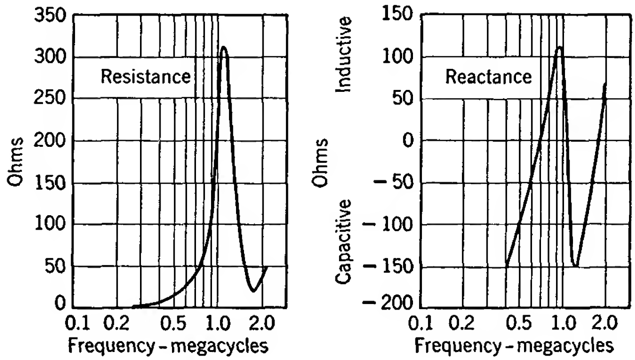


FIG. 29. Variations of driving-point input impedance for a vertical amplitude-modulation broadcast antenna. The tower is 350 feet high. The height of the mounting base and base insulators is 14 feet. The antenna ground system consists of buried conductors extending radially. Impedance measurements made between tower and ground system. (Courtesy General Radio Co. and Radio Station WEEI.)

35 ohms. This is the **natural frequency,** defined[34] as the "lowest resonant frequency obtained without added inductance or capacitance." At 0.685 megacycles, the antenna functions as a grounded

quarter-wave antenna with voltage and current distribution as shown in Fig. 30(*a*). At 1.070 megacycles, the driving point input impedance is 310 ohms resistance, the antenna is operating in resonance as a half-wave antenna fed at the end, and the voltage and current distribution are as shown in Fig. 30(*b*). At 1.715 megacycles the antenna of Fig. 29 is resonant, operation is as a three-quarter wave antenna, and the impedance is 20 ohms resistance. The voltage and current are distributed as in Fig. 30(*c*). An antenna does not have to be driven at a resonant frequency; however, the frequency at which it is driven affects the shape of the radiation pattern.

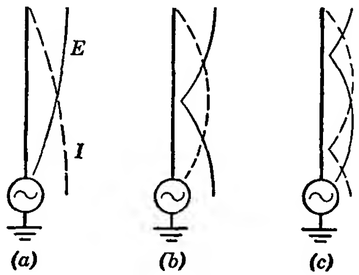


FIG. 30. Voltage and current distributions for a vertical antenna when driven at various frequencies so that it functions (*a*) as a quarter-wave antenna, (*b*) as a half-wave antenna, and (*c*) as a three-quarter wave antenna.

Other Periodic Antennas. The so-called **resonant Vee antenna** is of interest. One element consists of two wires arranged in a V, with the open end directed toward the distant receiving station. Several V elements may be arranged one above the other; also, V elements may be used as reflectors to give radiation in one direction only.[36]

Aperiodic (Non-resonant) Antennas. These antennas in their usual form are directional and are used for point-to-point service. **Aperiodic antennas** are defined[34] as "those designed to have constant impedance over a wide range of frequencies due to the suppression of reflection within the antenna system. These include terminated wave antennas and terminated rhombic antennas."

Because aperiodic antennas operate without wave reflection, they do not have standing (or stationary) waves on them. In a sense aperiodic antennas are similar to terminated transmission lines.

The Wave Antenna. This is defined[34] as "a directional antenna composed of a system of parallel, horizontal conductors from one-half to several wavelengths long, and terminated to ground at the far end in its characteristic impedance." The wave antenna, or **Beverage antenna**[34] has been extensively used as a long-wave, low-frequency, antenna. In its simplest form it consists of a single horizontal wire about one wavelength long pointed toward the distant station. The end of the line nearer the distant transmitting station is grounded through an impedance approximately equal to the characteristic im-

pedance of the antenna when regarded as a transmission line consisting of one wire and ground return. The other end is connected through the receiving equipment to ground.

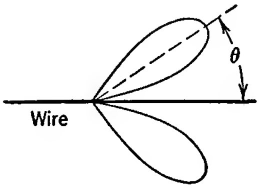


FIG. 31. The free-space radiation pattern for a lossless wire several wavelengths long, and terminated so that there are no standing waves, is as shown. The angle θ at which radiation is maximum is as given by Fig. 32.

The action of this antenna is explained[52] as follows: The incoming wave front is tilted forward slightly; thus, when an incoming wave travels *along* the antenna, a voltage is induced in it.[52] This voltage increases as the wave nears the receiving station. A wave along the line from the *opposite* direction will be absorbed at the terminating impedance. Waves coming from broadside will induce very little voltage, and thus the wave antenna is highly directive.

The Rhombic Antenna.[53, 54, 55] The rhombic antenna is extensively used in both transmitting and receiving in short-wave, high-frequency, radio telephone and telegraph systems for point-to-point communication. It is defined[34] as "an antenna composed of long-wire radiators comprising the sides of a rhombus." These conductors often are several wavelengths long and are terminated in an impedance at the "distant" end much like a transmission line so that no wave reflection exists on them. To determine the radiation pattern from a rhombic antenna the radiation from a single terminated wire must first be investigated.

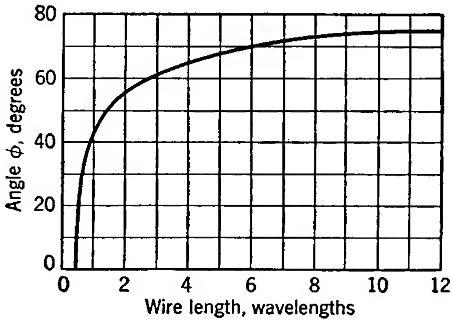


FIG. 32. For a lossless wire several wavelengths long, and in free space, maximum radiation occurs at the angle $\theta = (90° - \phi)$.

The radiation pattern from a long wire in free space is[6]

$$E = \frac{60I_0}{d} \cot \frac{\theta}{2} \sin \frac{pL}{2}, \quad (17)$$

where E is the field strength in volts per meter, I_0 is current input to the wire in amperes, d is the distance in meters, θ is the angle with respect to the axis of the wire, L is the wire length in wavelengths, and $p = (2\pi/\lambda)(1 - \cos \theta)$. Neglecting the minor lobes, Fig. 31 gives the free-space radiation pattern. The angle θ can be found

from Fig. 32, where $\theta = (90 - \phi)$. This curve shows that the angle of maximum radiation is almost independent of the length of the wire

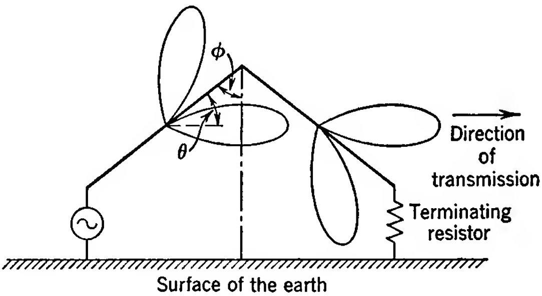


FIG. 33. The two wires of the tilted-wire antenna are held above the earth by a vertical pole at the center. For the proper height, the two lobes add to produce a strong signal in the direction of transmission. The actual radiation pattern would be affected considerably by reflection from the earth. The terminating resistor prevents wave reflection, giving a non-resonant antenna.

for wire lengths of over several wavelengths, a very important point as will be explained later. Before continuing the discussion of the rhombic antenna, two other antennas should be considered.

A **tilted-wire antenna** can be arranged as in Fig. 33. This simple structure is useful at the upper end of the high-frequency band. By the proper selection of wire length and angle ϕ, the antenna can be designed[37] so that two of the lobes combine for directional transmission and so that the other two largely cancel.

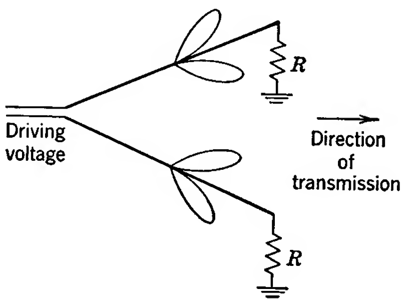


FIG. 34. The non-resonant Vee antenna is composed of two wires several wavelengths long, and connected to ground through terminating resistors. If the wires are arranged at the proper angles, the signals radiated, as indicated by the lobes, add, giving a strong signal in the desired direction. The plane of the two wires is parallel to the surface of the earth. The actual radiation pattern is affected by the earth.

A **non-resonant Vee antenna** is arranged as shown in Fig. 34. The two wires are horizontal above the earth. For the proper wire length and angles, a highly directional antenna is obtained. The wires are several wavelengths long and terminated so that no reflection occurs.

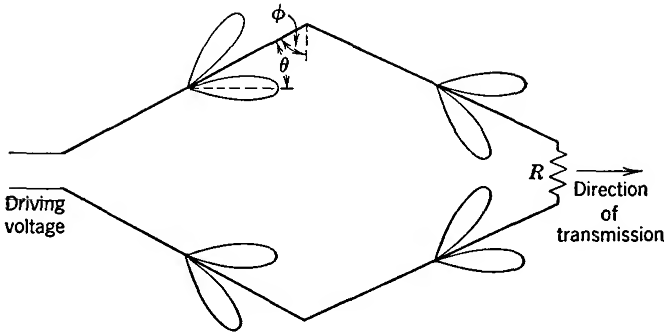
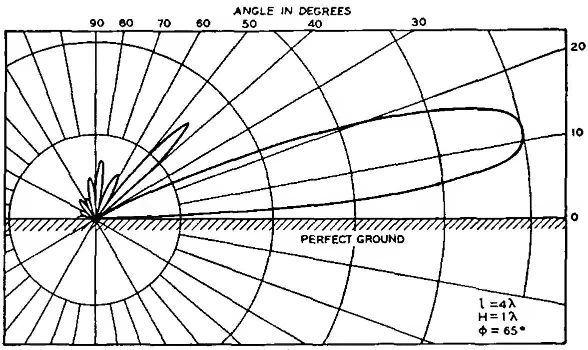
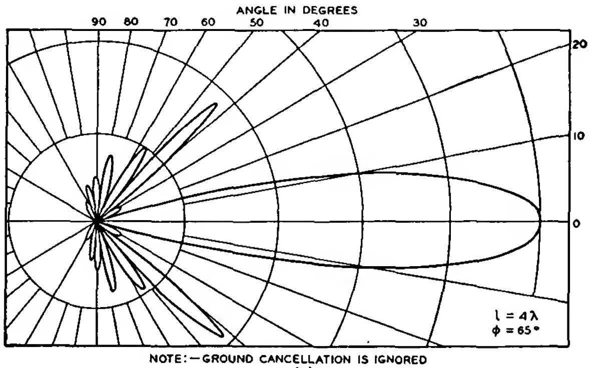


FIG. 35. The non-resonant rhombic antenna is composed of four wires each several wavelengths long and arranged as indicated. A terminating resistor R is used to prevent reflection. The radiation from each wire is as shown by the lobes and the individual radiations combine to give the directional patterns of Fig. 36. This is the top view of the structure. It is supported in a plane parallel to the surface of the earth.



(a)



(b)

FIG. 36. As shown in the upper illustration, the horizontal rhombic antenna having the specifications noted is most sensitive in reception (and also in transmission) to a wave or ray coming at an angle of about 10° with the horizontal. As the lower illustration indicates, the antenna is highly directional in the horizontal plane. (Reference 54.)

Horizontal rhombic antennas are shown in Fig. 35, and also in Fig. 20 of Chapter 13. This is one of the most satisfactory of all types of directional antennas. The relative radiations in vertical and horizontal planes are shown in Fig. 36 for a typical rhombic antenna composed of wires each 4λ in length, held horizontally one wavelength above the earth, and with $\phi = 65°$ (see Fig. 32). The radiation of each wire, as indicated by the lobes in Fig. 35, combine to give the patterns of Fig. 36.

The horizontal rhombic antenna requires no connection to ground, is non-resonant, and will operate well over a considerable band of frequencies because the angle ϕ of Fig. 32 varies but little with wavelength. This antenna is low in first cost, easy to maintain, and can be designed to meet a variety of operating conditions. It is important to note that it sends and receives best at some vertical angle with the surface of the earth. This angle can be varied by proper design. This adapts the horizontal rhombic antenna to sky-wave transmission and reception at high frequencies.

REFERENCES

1. *Standards on Radio Wave Propagation. Definition of Terms.* I.R.E. 1942.
2. McIlwain, K., and J. G. Brainerd. *High Frequency Alternating Currents.* Second Edition. John Wiley and Sons.
3. Skilling, H. H. *Fundamentals of Electric Waves.* John Wiley and Sons.
4. King, R. W. P., H. R. Mimno, and A. H. Wing. *Transmission Lines Antennas and Wave Guides.* McGraw-Hill Book Co.
5. *Electrical Engineering,* May, 1946, Vol. 65, No. 5, p. 238.
6. Terman, F. E. *Radio Engineers' Handbook.* McGraw-Hill Book Co.
7. Colwell, R. C. The Troposphere and Radio Waves. *Proc. I.R.E.,* July, 1940, Vol. 28, No. 7.
8. Dellinger, J. H. The Role of the Ionosphere in Radio-wave Propagation. *Trans. A.I.E.E.,* Supplement, 1939, Vol. 58.
9. Darrow, K. K. Analysis of the Ionosphere. *Bell System Tech. J.,* July, 1940, Vol. 19, No. 3.
10. Baker, W. G., and C. W. Rice. Refraction of Short Radio Waves in the Upper Atmosphere. *Trans. A.I.E.E.,* Vol. 45, 1926.
11. Rice, C. W. Short-wave Radio Transmission and its Practical Uses. *QST,* July, 1927, Vol. 12, No. 7.
12. Heising, R. A. Ionized Regions in the Atmosphere. *Bell Lab. Record,* February, 1928, Vol. 5, No. 6.
13. Austin, L. W. Solar Activity and Radiotelegraphy. *Proc. I.R.E.,* February, 1932, Vol. 20, No. 2.
14. Anderson, C. N. Correlation of Long Wave Transatlantic Radio Transmission with Other Factors Affected by Solar Activity. *Proc. I.R.E.,* March, 1928, Vol. 16, No. 3.
15. Schafer, J. P., and W. M. Goodall. Kennelly-Heaviside Layer Studies Employing a Rapid Method of Virtual-Height Determination. *Proc. I.R.E.,* July, 1932, Vol. 20, No. 7.

16. *Standards of Good Engineering Practice Concerning Standard Broadcast Stations (550–1600kc.).* Federal Communications Commission.
17. Feldman, C. B. The Optical Behavior of the Ground for Short Radio Waves. *Proc. I.R.E.*, June, 1933, Vol. 21, No. 6.
18. Burrows, C. R. Radio Propagation over Plane Earth—Field Strength Curves. *Bell System Tech. J.*, January, 1937, Vol. 16, No. 1.
19. Burrows, C. R., and M. C. Gray. The Effect of the Earth's Curvature on Ground-wave Propagation. *Proc. I.R.E.*, January, 1941, Vol. 29, No. 2.
20. Norton, K. A. The Calculation of Ground-wave Field Intensity Over a Finitely Conducting Spherical Earth. *Proc. I.R.E.*, December, 1941, Vol. 29, No. 12.
21. Terman, F. E. *Radio Engineering.* McGraw-Hill Book Co.
22. Report of Committee on Radio Wave Propagation. *Proc. I.R.E.*, October, 1938, Vol. 26, No. 10.
23. *Reference Data for Radio Engineers,* second edition. Federal Telephone and Radio Corp.
24. Potter, R. K. Transmission Characteristics of a Short-wave Telephone Circuit. *Proc. I.R.E.*, April, 1930, Vol. 18, No. 4.
25. Burrows, C. R. The Propagation of Short Waves over the North Atlantic. *Proc. I.R.E.*, September, 1931, Vol. 19, No. 9.
26. Prescott, M. L. The Diurnal and Seasonal Performance of High-frequency Radio Transmission over Various Long-distance Circuits. *Proc. I.R.E.*, November, 1930, Vol. 18, No. 11.
27. Burrows, C. R., and E. J. Howard. Short-wave Transmission to South America. *Proc. I.R.E.*, January, 1933, Vol. 21, No. 1.
28. Bullington, K. Radio Propagation at Frequencies above 30 megacycles. *Proc. I.R.E.*, October, 1947, Vol. 35, No. 10.
29. *Standards of Good Engineering Practice Concerning FM Broadcast Stations.* Federal Communications Commission.
30. Friis, H. T. Microwave Repeater Research. *Bell System Tech. J.*, April, 1948, Vol. 27, No. 2.
31. Pekeris, C. L. Wave Theoretical Interpretation of Propagation in Low-level Ocean Ducts, *Proc. I.R.E.*, May, 1947, Vol. 35, No. 5.
32. Robertson, S. D., and A. P. King. The Effect of Rain upon the Propagation of Waves in the 1- and 3-Centimeter Regions. *Proc. I.R.E.*, April, 1946, Vol. 34, No. 4.
33. Mueller, G. E. Propagation of 6-Millimeter Waves. *Proc. I.R.E.*, April, 1946, Vol. 34, No. 4.
34. *Standards on Transmitters and Antennas. Definitions of Terms.* I.R.E., 1942. *Standards on Antennas, Modulation Systems, Transmitters. Definitions of Terms.* I.R.E., 1948.
35. *American Standard Definitions of Electrical Terms.* A.I.E.E., 1941.
36. Carter, P. S., C. W. Hansell, and N. E. Lindenblad. Development of Directive Transmitting Antennas by R.C.A. Communications, Inc. *Proc. I.R.E.*, October, 1931, Vol. 19, No. 10.
37. *The A.R.R.L. Antenna Book.* American Radio Relay League.
38. Oswald, A. A. Transoceanic Radio Equipment, Short-wave Transmission. *Trans. A.I.E.E.*, April, 1930, Vol. 49, No. 4.
39. *Principles of Radar,* Staff, M.I.T. Radar School, McGraw-Hill Book Co.

40. Ballantine, S. On the Optimum Transmitting Wavelength for a Vertical Antenna over a Perfect Earth. *Proc. I.R.E.*, December, 1924. Vol. 12, No. 12.
41. Brown, G. H. A Critical Study of Broadcast Antennas as Affected by Antenna Current Distribution. *Proc. I.R.E.*, January, 1936, Vol. 24, No. 1.
42. Brown, G. H., R. F. Lewis, J. Epstein. Ground Systems as a Factor in Antenna Efficiency. *Proc. I.R.E.*, June, 1937, Vol. 25, No. 6.
43. Foster, R. M. Directive Diagrams of Antenna Arrays. *Bell System Tech. J.*, April, 1926, Vol. 5, No. 2.
44. Brown, G. H. Directional Antennas. *Proc. I.R.E.*, January, 1937, Vol. 25, No. 1.
45. Laport, L. A. Directional Antenna Design. *Electronics*, April, 1936, Vol. 9, No. 4.
46. Smith, C. E. Directional Antennas. *Cleveland Institute of Radio Electronics.*
47. Everest, F. A., W. S. Pritchett. Horizontal-Polar-Pattern Tracer for Directional Broadcast Antennas. *Proc. I.R.E.*, May, 1942, Vol. 30, No. 5.
48. Hutton, W. G., and R. M. Pierce. A Mechanical Calculator for Directional Antenna Patterns. *Proc. I.R.E.*, May, 1942, Vol. 30, No. 5.
49. Brown, G. H., and W. C. Morrison. The RCA Antennalyzer—An Instrument Useful in the Design of Directional Antenna Systems. *Proc. I.R.E.*, December, 1946, Vol. 34, No. 12.
50. Smith, P. H. "Cloverleaf" Antenna for F.M. Broadcasting. *Proc. I.R.E.*, December, 1947, Vol. 35, No. 12.
51. Chamberlain, A. B., and W. B. Lodge. The Broadcast Antenna. *Proc. I.R.E.*, January, 1936, Vol. 24, No. 1.
52. Beverage, H., C. W. Rice, and E. W. Kellogg. The Wave Antenna. *Trans. A.I.E.E.*, Vol. 42, 1923.
53. Harper, A. E. *Rhombic Antenna Design.* D. Van Nostrand Co.
54. Bruce, E. Developments in Short-wave Directive Antennas. *Proc. I.R.E.*, August, 1931, Vol. 19, No. 7.
55. Bruce, E., A. C. Beck, and L. R. Lowry. Horizontal Rhombic Antennas. *Proc. I.R.E.*, January, 1935, Vol. 23, No. 1.

REVIEW QUESTIONS

1. Describe how radio energy is transmitted through space.
2. Distinguish between the induction field and the radiation field.
3. In what groups are radio frequencies sometimes classified?
4. What are the components of the ground wave?
5. Explain what is meant by diffraction, refraction, and absorption of radio waves.
6. What part does the ionosphere play in radio transmission?
7. What part does the troposphere play in radio transmission?
8. Describe how radio waves are reflected by the ionosphere.
9. What important ionospheric layers exist in summer? In winter? Does the E layer disappear completely at night?
10. What factors seem to influence the ionosphere?
11. What is virtual height, and how is it measured?
12. What is meant by penetration frequency? By maximum usable frequency?
13. What are the ordinary and extraordinary rays?
14. What is skip distance? Does the ground wave enter into determining it?

15. Discuss the phenomena that occur when radio waves strike the earth.
16. Define sky wave, ground wave, direct wave, and ground-reflected wave.
17. A vertically polarized wave is being propagated along the surface of the earth. Why must the wave front tilt forward slightly?
18. By what two paths does radio energy reach a distant receiving antenna?
19. What is fading, and how is it caused?
20. Describe direct-wave propagation.
21. Of what frequency are radio waves that are greatly affected by rain?
22. Distinguish between periodic and aperiodic antennas. By what names are they usually called?
23. Can the reactive component of the input impedance of an antenna be zero? Can the resistive component be zero?
24. At what frequencies is an antenna resonant? At all such frequencies is the input impedance the same value?
25. What important assumptions are made in determining ground-reflection factors?
26. Explain in simple terms how to find the radiation pattern for an antenna near the earth.
27. Define broadside array, end-fire array, collinear array.
28. What will be the phase relations in the vertical members of Fig. 22?
29. Why does a straight vertical antenna 0.53λ in length make an excellent radiator?
30. Why are vertical antennas used in amplitude-modulation broadcast?
31. Why are directional antennas used in these systems?
32. How are antennas arranged to produce directional patterns?
33. Why are the conductors of a frequency-modulation broadcast antenna horizontal?
34. What is the difference between antenna resistance and radiation resistance?
35. The rhombic antenna is what basic type? How does it operate?

PROBLEMS

1. Use the data in Fig. 5 and compute the maximum usable frequency for communicating a given distance at several different times of day.
2. A grounded vertical quarter-wave antenna is fed 5.0 kilowatts at 550 kilocycles. Calculate the field strength produced at one mile, assuming no losses. Draw a circle showing the area covered by this station, if 0.5 millivolt per meter is the limit of satisfactory reception and if earth conditions are good. Repeat for a similar antenna fed the same amount of power at 1600 kilocycles. Draw a similar circle using the same center. Is there any relation between the areas of these circles and the frequency assignment of standard amplitude-modulation broadcast stations?
3. Use curves such as given in reference 16, and, for the conditions of Problem 2, plot two comparative circles for 550 kilocycles with maximum and minimum earth conductivities, and plot two comparative circles for 1600 kilocycles for maximum and minimum conductivities.
4. If two ships have antennas mounted 150 feet above the water line, how far can they communicate by direct-ray transmission?
5. Draw diagrams such as shown in Fig. 12 for a $\frac{3}{4}\lambda$ antenna, and a 1.0λ antenna. What will be the driving-point impedances?

6. Calculate the data for and plot the ground-wave reflection diagram for a horizontal half-wave antenna $\frac{3}{4}\lambda$ above the earth.
7. Draw arrows for instantaneous current flow on the antenna of Fig. 22, proving that it has the directional properties explained.
8. Calculate the horizontal radiation pattern for the antenna array considered on page 473, except that the spacing is 120° and the current ratio is 0.8.
9. Repeat the calculations on page 476 using equation 16 for an azimuth angle of 45° and a zenith angle of 45°.
10. A broadcast antenna has an input impedance of 37 ohms resistance. If it is drawing 1.0 kilowatt, what is the current input, and the voltage across the base insulator? Repeat for 5, 10, and 50 kilowatts.
11. If each sloping wire of Fig. 33 is 2λ in length, what must be the height in wavelengths? What would be the height in feet at 10 megacycles? At 30 megacycles?
12. If each of the four wires of the rhombic antenna of Fig. 35 is 2λ in length and if maximum directivity in the horizontal plane is the only objective, determine the important angles and dimensions for the antenna of Fig. 35.

CHAPTER 13

RADIO SYSTEMS

Introduction. Radio systems are operated by incoming signals of several types. In **radio telephony,** speech signals control the transmitter output. In **radio telegraphy,** the signals are code characters. In **radio broadcasting,** the signals are those of speech and music. In **picture transmission, facsimile transmission,** and **television,** the information to be transmitted is in the form of electric-current variations produced as the image to be transmitted is scanned.

In a radio system, signals such as just discussed are used to **modulate**[1] a radio-frequency carrier, and the information to be transmitted is translated to a frequency band in the radio-frequency spectrum (page 442). The information, then existing at radio frequencies, is transmitted through space and at the receiving station is **demodulated**[1] and returned to its original form for reception.

Modulation. This is the process whereby the information in the modulating low-frequency signal wave is translated to the desired radio-frequency band (page 414). By the use of different assigned carrier frequencies, the information to be transmitted by various radio transmitters can be raised to different locations in the radio spectrum. Thus, many radio signals can be transmitted simultaneously through space, and the desired signal can be selected on a frequency basis at the receiving station.

Many types of modulation are possible, and several different systems are in use. These sometimes are classified[2] as follows: **amplitude modulation; angle modulation** (which includes frequency modulation and phase modulation); and **pulse modulation.**

Amplitude Modulation or AM. This system of modulation has been used since the beginning of radio and, of all systems, is most extensively used today. It is defined[2] as "modulation in which the amplitude of a wave is the characteristic subject to variation." An **amplitude-modulated wave** is defined[2] as follows: "An amplitude-modulated sinusoidal wave is one whose envelope contains a component similar to the wave form of the signal to be transmitted." Amplitude-modulated waves were discussed in Chapter 11. By the process of amplitude modulation, two sidebands are created, and *each* of these contains, independently of the other, all the information of

the original modulating signal to be transmitted. These sidebands are created by simultaneously impressing the carrier wave and the modulating signal wave on a modulating circuit that distorts the waves and produces, because of the distortion, sum and difference frequencies. This has been summarized by Everitt who, in discussing *amplitude* modulation, has written,[3] "Modulation is obtained whenever a component of the output wave is made proportional to the *product* of two input waves."

If a radio-program signal, which covers an audio-frequency band of approximately 100 to 5000 cycles, is used to modulate a carrier-frequency wave of 1,000,000 cycles, the modulated wave, as transmitted by a standard amplitude-modulation broadcast station, will consist of three components: the carrier wave of 1,000,000 cycles; the upper sideband extending from 1,000,100 to 1,005,000 cycles; and the lower sideband extending from 999,900 to 995,000 cycles. As explained for carrier telephone systems, sometimes the carrier and one sideband are suppressed in radio systems, but this is not done (at present, 1949) in broadcasting.

Amplitude-modulated radio transmitters are of two general types, those employing high-level modulation, and those employing low-level modulation. **High-level modulation** is defined[2] as "modulation produced at a point in a system where the power level approximates that at the output of the system." **Low-level modulation** is defined[2] as "modulation produced at a point in a system where the power level is low compared to the power level at the output of the system."

Amplitude-Modulated Radio Transmitters.[2] Block diagrams of high-level and low-level radio transmitters are shown in Fig. 1.

High-Level Amplitude Modulation. A sinusoidal radio-frequency carrier wave is generated in the crystal-controlled oscillator (page 304), which may also include a buffer-amplifier stage. The carrier wave is amplified in a class C radio-frequency power amplifier (page 295) until it is of sufficient strength to be impressed on the modulator. The speech or program signal to be transmitted is impressed on the audio-frequency power amplifier, and the signal strength is increased until it is sufficient to be impressed on the modulator. As is evident in Fig. 1, the output of the modulator is impressed on the antenna. Thus, the power level at modulation approximates the output of the system, and the term high-level modulation applies.

Low-Level Amplitude Modulation. The carrier wave is generated as before and is then amplified, but not to so high a level as in the system just described. The speech or program signal to be transmitted is amplified in the audio-frequency power amplifier, but, again,

the power level is much less than with high-level modulation. After modulation, the carrier and sidebands are amplified in a class B radio-frequency power amplifier (page 295). This raises the power level to that required for radiation by the antenna (Fig. 1).

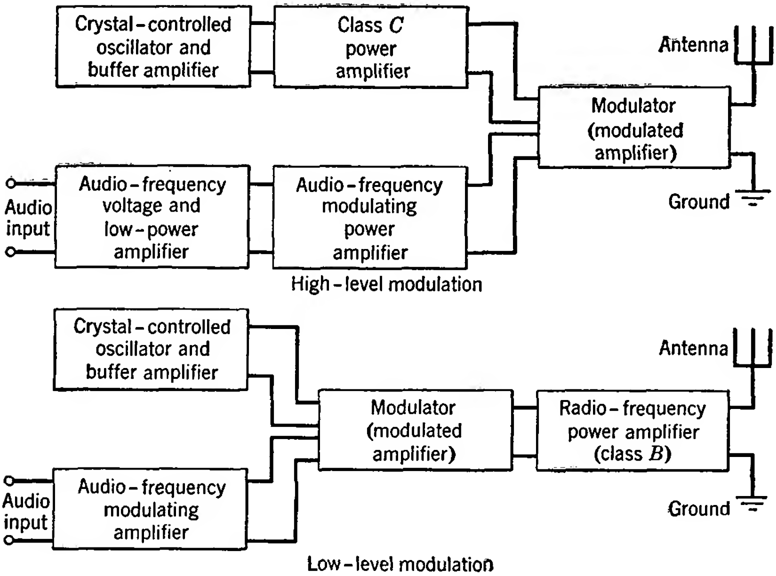


FIG. 1. Basic types of amplitude-modulation radio transmitters.

Methods of Amplitude Modulation. Many methods of amplitude modulation are used in radio transmitters, but, for most commercial broadcast and communication systems, only several types are employed. Unfortunately, in discussions of modulation, loose terminology exists. For instance, the audio-frequency amplifier of Fig. 1 is often called the modulator; a term that is more appropriate and is growing in use is **modulating amplifier,** because of the accepted concept that the audio wave modulates the carrier wave. The **modulator**[2] was defined on page 418 as "a device to effect the process of modulation." This is also called the **modulated amplifier,** defined[2] as "an amplifier stage in a transmitter in which the modulating signal is introduced and modulates the carrier."

Modulated Class C Amplifier, Modulating Wave Injected in Plate Circuit. A simplified circuit is shown in Fig. 2, and the equivalent circuit in Fig. 3. The magnitude of the instantaneous radio-frequency output voltage depends on the magnitude of the direct plate-to-cathode

voltage (page 295). For perfect modulation this relation should be linear. Then, as indicated in Fig. 3, when the audio voltage from the modulating amplifier varies in accordance with the signal to be transmitted, the plate voltage of the modulator, or modulated amplifier, will vary as shown by e_b of Fig. 3(*b*), and the plate current will vary as shown in (*c*). The parallel resonant circuit *C–L* is tuned to the carrier frequency (and sidebands), and the alternating current

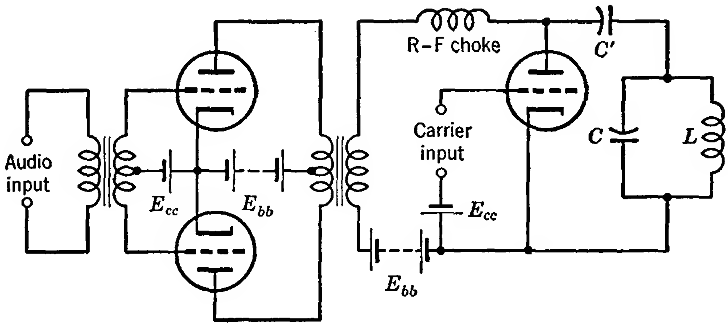


FIG. 2. Simplified circuit of a push-pull audio-frequency modulating amplifier (left) driving a modulated class C amplifier, or modulator, (right). The modulated output wave of Fig. 3 (*d*) appears in the circuit *L–C* which is coupled to the antenna in a high-level modulation circuit. The neutralizing circuit (page 299) is omitted.

through coil *L* and the alternating voltage across the L–C circuit will vary as in (*d*). This is an amplitude-modulated wave (page 412), and this output circuit is coupled to the antenna, the arrangement shown being for high-level modulation.

For conditions of operation commonly used, the peak value of the carrier voltage impressed on the grid of the modulated class C amplifier drives the control grid positive. With the grid bias value known (page 295) the necessary carrier voltage can be estimated. Since the plate voltage varies at an audio signal rate, the cutoff point varies and the carrier power input varies. For a triode, the carrier power input required is approximately 10 per cent of the modulated plate-power output. The audio-frequency modulating power amplifier of Fig. 2 is often class B push-pull (page 294). In amplitude modulation, the magnitude of the power output of the *carrier-frequency component* is the same before and during modulation. The power that goes into the carrier wave output of the modulated amplifier comes from the plate power supply E_{bb}. The power that goes into the sidebands comes from the audio-frequency modulating amplifier. For 100 per cent

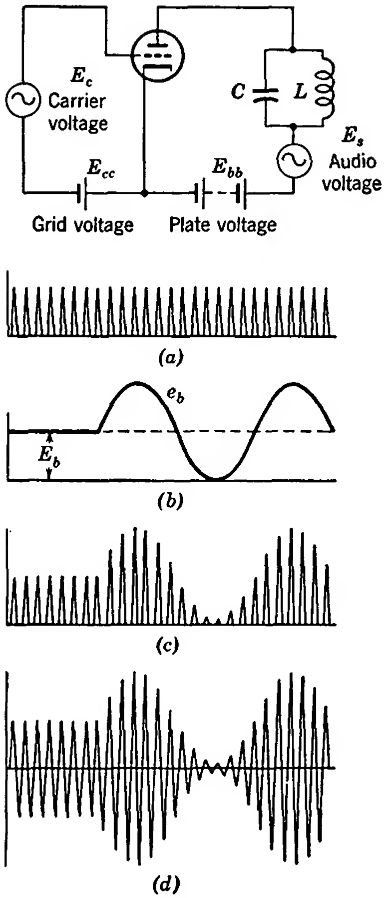


FIG. 3. Equivalent circuit for a modulated class C amplifier, and curves showing the operation. (a) Plate current with carrier only impressed. (b) Direct voltage E_b on the plate, and total voltage e_b when modulating signal is impressed. (c) Plate current when carrier voltage and modulating voltage both are impressed. (d) Voltage drop across tuned parallel circuit and current in tuning coil and condenser.

modulation, the power in the two sidebands is 50 per cent of that in the carrier.

Modulated Class C Amplifier, Modulating Wave Injected in Grid Circuit. A simplified circuit is shown in Fig. 4(*a*) which includes a neutralizing capacitor C_N. The grid is biased considerably beyond cutoff as Fig. 4(*b*) indicates. With the carrier voltage only impressed, the instantaneous grid voltage and plate current are as shown by *A–B*. When the modulating voltage E_s from the modulating amplifier is impressed in series with the carrier voltage, the instantaneous grid voltage and plate current are as shown in *B–C*. When this current flows through the tuned circuit at the right in Fig. 4(*a*), the alternating voltage across this circuit and the alternating current through the inductor will be as shown by Fig. 3(*d*). Usually the antenna or the feeder is inductively coupled to this inductor.

The grid of the tube in Fig. 4 is driven positive; therefore, the class C amplifier, associated with the crystal in the lower diagram of Fig. 1, and the audio-frequency modulating amplifier (of Fig. 1) must have power output capacities sufficient to supply the losses of the grid circuit of the modulated class C tube. But the audio-modulating amplifier does not have to supply the power for the sidebands. Each amplifier must supply about 10 per cent of the modulated output power. The circuit of Fig. 4 is for low-level modulation (Fig. 1).

Other Methods of Amplitude Modulation. The balanced modulator of page 421 has been used in small radio broadcast transmitters.[4] The

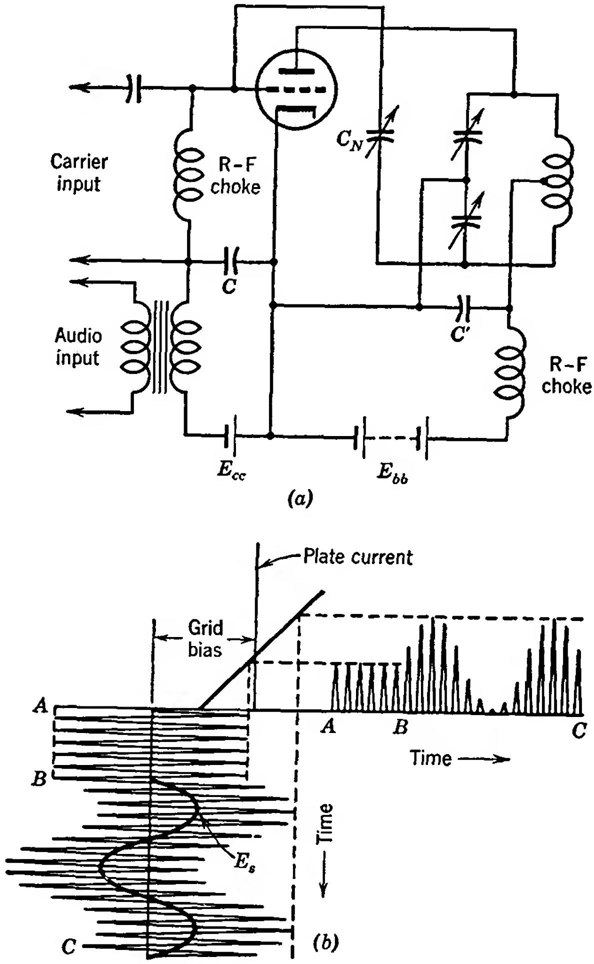


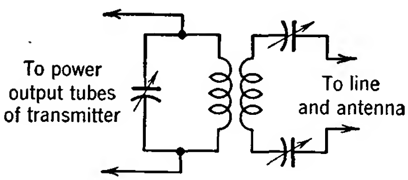
FIG. 4. Simplified circuit (*a*) for a modulated class C amplifier with modulating wave injected in the grid circuit. This is often called grid-bias modulation, or grid modulation, and is commonly used in low-level modulation. In (*b*) is shown how this modulator operates. When both the carrier and the modulating signal E_s are injected, the instantaneous grid-voltage variations and the instantaneous plate-current variations are shown by *B–C*. These plate-current changes produce the conventional amplitude-modulated wave in the plate circuit.

circuit was arranged to suppress the audio signal and transmit the carrier component. Another circuit sometimes used is called *cathode*

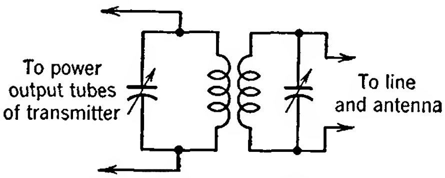
modulation in which the carrier wave is impressed in the grid circuit, and the modulating wave simultaneously is impressed between cathode and ground.

The tubes indicated in the modulated amplifiers of Figs. 2 and 4 are triodes, but beam-power tubes or pentodes are often used for this purpose. When pentodes are used, **suppressor-grid modulation** is possible. With this arrangement the carrier wave is impressed, or injected, in the control-grid circuit, and the audio-modulating wave is injected between the suppressor grid and cathode.[5]

Antenna Feeder Systems and Impedance-Matching Networks. Sometimes a radio transmitter is connected directly to the radio antenna, but often the transmitter and antenna are some distance apart, and the radio-frequency power from the transmitter is fed to the antenna over either an open-wire transmission line (Chapter 6) or a coaxial cable (Chapter 7). Sometimes the antenna feeder systems (transmission lines and cables) *are not* terminated in their characteristic impedances so they *have* standing waves on them (called **mismatched,** or **resonant, feeders**). Sometimes the feeders *are* terminated in their characteristic impedances and *do not have* standing waves on them (called **matched,** or **non-resonant, feeders**).



(a) For coupling a transmitter at a low-impedance point



(b) For coupling a transmitter at a high-impedance point

FIG. 5. Circuits sometimes used in connecting mismatched, or resonant, antenna feeders to the output circuit of a radio transmitter, so that the output power amplifier will be loaded correctly and power will flow from the transmitter to the feeder.

Mismatched, or Resonant, Feeders. Assume, for the moment, that a 500-ohm open-wire line is to drive a horizontal half-wave antenna at the center where the input impedance is approximately 73 ohms. If the line is directly connected to the antenna it will be mismatched and standing waves will exist on the line. At the transmitter the impedance of the line (with the antenna connected) must be matched to the output circuit of the transmitter so that power will flow into line and antenna.

The input impedance of the line terminated with the antenna will depend on the frequency, physical length of the line, characteristic

impedance of the line, and the impedance of the antenna. Two common coupling circuits are shown in Fig. 5. These impedance-transforming circuits can be designed in accordance with the theory in Chapter 3. There are advantages to operating with a mismatch at the antenna. If a radio system operates at several frequencies that are changed frequently and if the line is mismatched at the antenna, then adjustments can be made at the transmitter and no changes need be made in antenna matching networks. On the other hand, radiation from the feeders is slightly greater, and the overall efficiency is lower; furthermore, the voltage is higher at certain points on the feeder, but usually this causes no insulation problems.

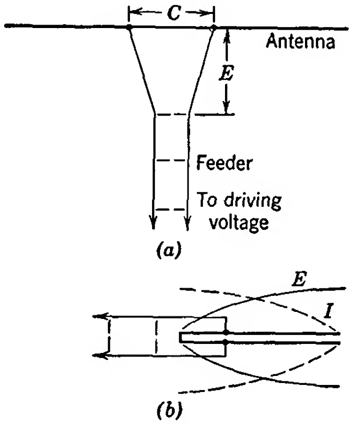


FIG. 6. In (*a*) is shown a method of matching the impedance of a half-wave antenna to an open-wire transmission line. The dotted lines between the wires of the transmission line, or feeder, represent the positions of ceramic insulating spreaders. In (*b*) is shown a sketch indicating how different impedances are obtained by connecting at different points.

Matched, or Non-resonant, Feeders. If a radio transmission system is to operate at one frequency and maximum performance is desired, then an impedance transforming, or matching, network is often placed between the antenna and the feeder, and between the feeder and the transmitter.

The so-called **delta-matching transformer** of Fig. 6(*a*) is commonly used for connecting a half-wave antenna to a balanced transmission line. The dimensions

$$E = 123/f, \text{ and } C = 148/f \tag{1}$$

are used.[6] The values of E and C are in feet, f is in megacycles, and the constants apply for a 600-ohm feeder line composed of two No. 12 A.W.G. copper wires 6 inches apart. The theory of this matching arrangement can be explained by referring to Fig. 6(*b*). If a half-wave antenna is "folded" into a quarter-wave line (to illustrate the theory), the current and voltage distribution will be as indicated. By moving the point of contact of the transmission line along the folded antenna, various ratios of voltage to current, and hence various impedances, are available. This partly explains the "C" portion of

Fig. 6(*a*). The other factor is that the "*E*" portion approximates an exponential transmission line (page 229) which has impedance-transforming properties.

What is known as **stub matching** is often used to match an antenna to a transmission line, particularly at very high and ultrahigh frequencies.[7] There are several ways of accomplishing this, one of which is shown in Fig. 7. The impedance of the stub is in parallel with the impedance of the feeder and antenna to the right of the stub. Such a combination will have impedance-transforming properties,[8] and, by the proper length and location of the stub, an impedance match is obtained, and no standing waves will exist on the feeder between the stub and the transmitter. The location and the length of the stub can be found from Fig. 8. The standing-wave ratio indicated on Fig. 8 is the ratio of voltage maximums to voltage minimums, or current maximums to current minimums, before the stub is attached. These can be measured in several ways, simple methods being as follows: For *measuring voltage ratios*, attach a low-impedance thermomilliammeter to the end of a quarter-wave section of transmission line. The input impedance of this combination will be very high, and it can be moved along the line as a voltmeter, the indications of the thermomilliammeter being an indication of the voltages along the line. For *measuring current ratios,* attach two short sturdy wire hooks directly to the thermomil-

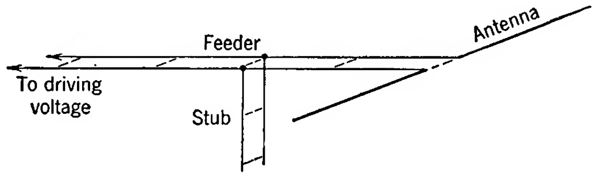


FIG. 7. A stub is often used to match an antenna to a transmission line, or feeder, so that standing waves do not exist between the stub and the transmitter.

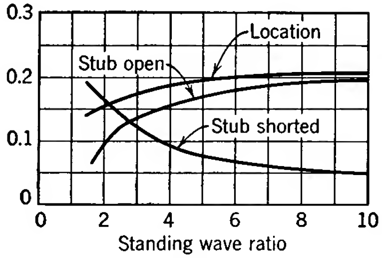


FIG. 8. Curves for determining the length and the location of stubs for matching an antenna to a transmission line, or feeder. The distant end of the stub may be open or shorted. The "location" curve gives the location on the line from a voltage maximum toward the transmitter if a shorted stub is used, and from a voltage maximum toward the antenna if an open stub is used. The two other curves give the proper stub length, all dimensions being in terms of the wavelength, λ. The characteristic impedance of the stub is assumed to equal that of the line.

liammeter so that it can be hung from *one* wire. As the thermomilliammeter is moved along the wire, the indication will be a measure of the current in the wire. The instrument used should read effective values, or the readings must be corrected.

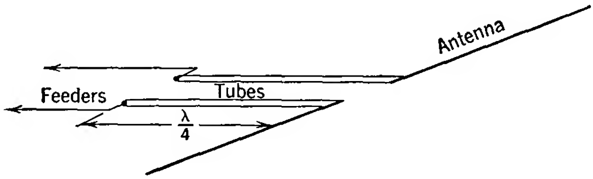


FIG. 9. Two large wires or tubes placed close together can be used to match an antenna to a transmission line, or feeder.

What is known as a **reentrant network**[9] is also used for impedance matching. With this arrangement a short loop of wire extends as a loop from one point on each feeder wire to another point on the same feeder wire.

Another matching device is the **quarter-wave matching section** of Fig. 9. Note that the two large wires or tubes are exactly $\frac{1}{4}\lambda$ in length. This is because the input impedance of a half-wave antenna (with which this system is used) is 73-ohms resistance, and the characteristic impedance of radio-frequency transmission-line feeders also is pure resistance. The characteristic impedance Z_{OS} of the quarter-wave matching section to be used is[10]

$$Z_{OS} = \sqrt{Z_{OL}Z_A}\,, \tag{2}$$

where Z_{OL} is the characteristic impedance of the transmission line and Z_A is the driving-point, or input, antenna impedance; all values being in ohms. The size and spacing of the conductors for the quarter-wave matching section can be determined as explained in Chapter 6.

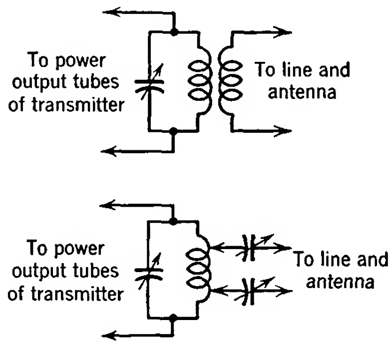


FIG. 10. Networks for connecting matched, or non-resonant, lines to radio transmitters.

Two circuits for connecting matched or non-resonant transmission-line feeders to a radio transmitter are shown in Fig. 10. These can be designed in accordance with discussions in Chapter 3.

For *coaxial cables feeding broadcast antennas* the two impedance-matching circuits of Fig. 11 are used. In (*a*), the antenna is an ungrounded vertical tower with a base insulator and a **ground system** composed of buried radial wires. This matching circuit can be designed as explained in Chapter 3. In (*b*), the antenna is a grounded vertical tower, the ground being buried radial wires as before. This impedance-matching circuit, and in fact the entire arrangement, will be recognized as one half of the circuit of Fig. 6. In the final adjustments a radio-frequency impedance bridge, instead of the coaxial cable, is connected to the end of the sloping feed wire, and the wire is moved up and down the tower until a resistance value equal to the characteristic impedance of the coaxial (usually 77 or 52 ohms) is found. Ordinarily, some inductive reactance is also measured, and this is neutralized with the capacitor of Fig. 11(*b*). At the transmitter, the coaxial cable would be connected to the last power output stage through an impedance-matching network such that the power output tube (or tubes) operates into the proper load impedance. In broadcast installations, phase-shifting networks[11] are used at the transmitter if directional antennas are employed. These networks produce the proper current ratios and phase shifts to give the desired directional patterns (page 473).

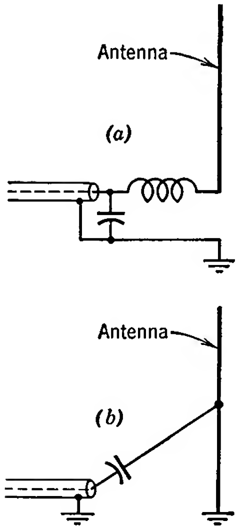


FIG. 11. Methods of connecting coaxial cables to radio-broadcast antennas. Diagram (*a*) is for a base-insulated antenna, and (*b*) is for a grounded antenna.

Demodulation (Detection) of Amplitude-Modulated Waves.[1, 2] In amplitude *modulation* the message to be transmitted and a carrier wave are impressed on a circuit that distorts and produces sum and difference frequencies, or sidebands, which contain the information to be transmitted. By the same process, if the carrier and sidebands are simultaneously impressed on a circuit that distorts, sum and difference frequencies are created. The difference frequencies are the desired demodulated signal and exist at audio frequencies. **Demodulation** was defined[1] on page 421. **Detection** is[2] the "process by which a wave corresponding to the modulating wave is obtained in response to a modulated wave." Fundamentally, demodulation and detection are the same process as modulation, a fact stressed in Chapter 11.

Amplitude-Modulation Radio Receivers.[12] Two basic types are used, the **tuned-radio-frequency** receiver and the **superheterodyne receiver.** The first of these is not used extensively.

The Tuned-Radio-Frequency Receiver. This is shown in the block diagram of Fig. 12. The signal voltage from the antenna is impressed on several stages of tuned-radio-frequency voltage amplifiers that increase the magnitude of the carrier and the two sidebands constituting the received message or program signal. This signal is then impressed on the demodulator, or detector, where it is distorted, and sum and difference frequencies are produced. If the carrier and the sidebands are as listed on page 489, the sum frequencies are in the vicinity of 2

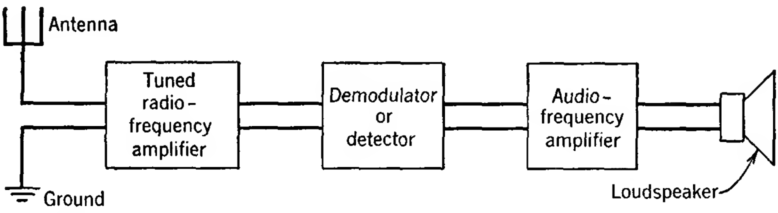


FIG. 12. Block diagram of a tuned-radio-frequency radio-receiving set.

megacycles, and the difference frequencies will constitute two bands from 100 to 5000 cycles. These two bands are in phase and combine to produce the desired audible output signal voltage. This is passed by the audio-frequency amplifier, which incorporates both voltage and power amplification, and the output drives the loudspeaker (or headphones) reproducing the message or program transmitted.

The demodulator, or detector, may be one of several types. It is possible to use vacuum-tube plate-circuit demodulation similar to that used in early carrier telephone systems (page 421). However, in the tuned-radio-frequency receivers that once were used, the distortion required for demodulation was produced in the grid circuit of the tube. The detector tube was operated with negligible grid bias, and, when the amplified received signal composed of the carrier and the sidebands was impressed, distorted grid current flowed to cathode. A parallel circuit composed of a resistor and capacitor was placed in series with the grid. The circuit constants were such that, when the distorted grid current flowed through it, the difference frequencies (that is, the desired audio signal) caused an audio-frequency voltage drop across the resistor-capacitor combination, but radio-frequency components caused negligible drop. This audio-voltage drop, being between the grid and cathode, was amplified in the plate circuit of

the detector tube and, after additional amplification, drove the loudspeaker. This circuit was called the **grid-leak detector.**

A tuned-radio-frequency receiving set is simple and has advantages for some purposes such as fixed-frequency communication. For broadcast reception in the standard amplitude-modulation range, it must tune from 550 to 1600 kilocycles, and this tuning offers design problems. The set is not very selective in its usual form, and the early type of detection caused serious distortion on signals with high percentage modulation.

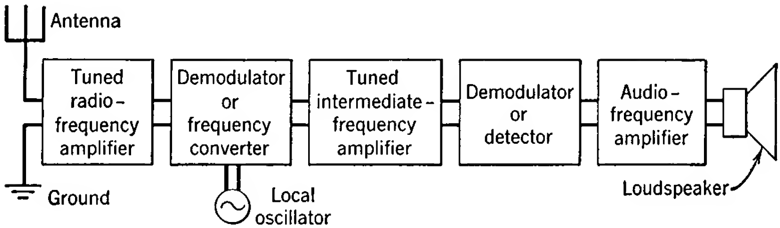


FIG. 13. Block diagram of a superheterodyne radio-receiving set.

The Superheterodyne Radio Receiver. As previously mentioned, the superheterodyne is almost universally used. The name means little from the standpoint of the electrical principle of operation. It has been called a **double-detection receiver.**

A block diagram of the superheterodyne radio receiver is shown in Fig. 13. The functioning of a tuned-radio-frequency amplifier has been explained. The demodulator (also called **first detector, converter,** or **mixer**) operates as follows: Suppose that the amplitude-modulated signal of page 489, composed of a one-megacycle carrier and two sidebands, has been received and is impressed between grid and cathode of the frequency converter. Also, at this same time assume that a single frequency of 1,455,000 cycles is impressed between grid and cathode of the converter, which is a vacuum tube operated so that non-linear distortion occurs. Sum and difference frequencies will be produced by this tube, and among the several new frequencies there will exist the components of 455,000 cycles, 455,100 to 460,000 cycles, and 454,900 to 450,000 cycles. These are the carrier and the two sidebands translated, or moved down, to a new location in the radio spectrum. All the information of the original modulating signal now exists at these new frequencies; hence the term *frequency converter*. It should be remembered, however, that the basic principle involved is the same as that of the usual modulating circuit; that is,

one of non-linear distortion and the creation of sum and difference frequencies.

The high selectivity of the superheterodyne is achieved through the selective action of the **intermediate-frequency amplifier.** This is a carefully tuned radio-frequency amplifier (page 290) of high voltage gain, that is adjusted to amplify and pass a band of about 450,000 to 460,000 cycles only, and to offer much attenuation to other frequencies. This amplifier will amplify the carrier and sidebands at the "intermediate frequencies" of the preceding paragraph.

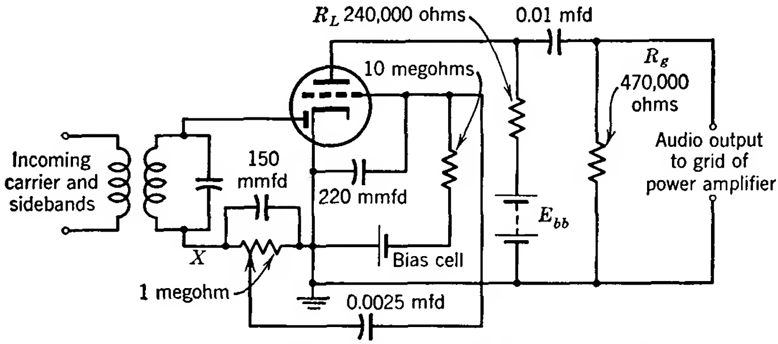
If an unwanted station exists at a carrier frequency of 1,020,000 cycles and if some of this signal is passed by the tuned-radio-frequency amplifier, this also will react with the 1,455,000-cycle signal from the local oscillator and produce a difference-frequency band centered at 475,000 cycles. This would be too high to pass through the intermediate-frequency amplifier. If it is desired to receive the 1,020,000-cycle station and reject the 1,000,000-cycle station, this can be accomplished merely by adjusting the local oscillator to 1,475,000 cycles.

The selected and amplified signal passing through the intermediate-frequency amplifier (usually called an **I–F amplifier**) is impressed on another demodulator, or **second detector,** as it is usually called. In this, sum and difference frequencies are produced from the so-called intermediate frequencies. That is, sum frequencies of about 1,000,000 cycles will be produced, and difference frequencies of from 100 to 5000 cycles will be produced. The difference frequencies are the desired audio-frequency band containing the information or program transmitted from the distant station.

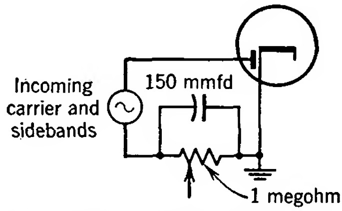
The **diode detector** shown in Fig. 14(*a*) is the demodulator or second detector in the superheterodyne radio receiver. The tube shown is a combination diode and triode. The diode portion is used for demodulation, and the triode portion for amplification. These two functions are shown separately in (*b*) and (*c*).

The carrier and two sidebands, existing at the intermediate frequencies, are simultaneously distorted in the diode rectifying portion (*b*), and the desired audio-frequency component, containing the transmitted message or program, is created. This will flow through the circuit composed of the 150-micromicrofarad capacitor and 1.0-megohm resistor in parallel. The impedance of this combination is such that the audio-frequency components will cause considerable voltage drop, but the radio-frequency components will cause negligible voltage drop. A part of the audio-voltage drop across the 1.0-megohm resistor is selected by the manual volume control, indicated by the arrow head on this resistor, and this voltage is impressed between the

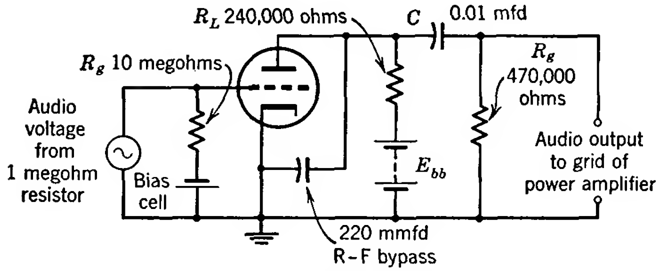
grid and cathode of the triode portion of the tube. The tubes often used have amplification factors of 100 for the triode portion, and con-



(a) Actual circuit of detector and audio voltage-amplifying stage.



(b) Rectifying or demodulating portion.



(c) Audio-frequency voltage-amplifying portion.

FIG. 14. Diode detector and audio-frequency voltage amplifier as used in a typical superheterodyne radio-receiving set for receiving amplitude-modulated signals.

siderable audio-frequency voltage amplification is produced in this stage.

An **automatic-volume-control voltage** is obtained between point X and ground in Fig. 14(a). Assume that the radio set is tuned to the carrier frequency of a station that is, for the moment, unmodulated.

The carrier will be rectified in the diode part of the tube, a direct current will flow through the 1.0-megohm resistor, and a direct voltage will exist across it, with the X terminal negative. If the radio signal transmission path improves, because of ionospheric or other changes, a stronger signal will be received, the voltage drop will increase, and point X will be more negative. If the radio signal path becomes less efficient, point X becomes less negative. The automatic-volume-control voltage is obtained between point X and ground and is the voltage just discussed. This is impressed as a grid bias on a **variable-mu** or **super-control** tube. This is usually one of the pentode voltage-amplifying tubes used in the intermediate-frequency amplifier, and it is constructed so that the amplification factor varies widely with changes in the grid-bias voltage. Thus, when the received signal is strong, the automatic-volume-control voltage used as a bias is large and the tube is operated so that its amplification is low. When the received signal is weak, the bias is low and the amplification produced by the tube is high. The preceding discussion was for an unmodulated carrier signal only. When a modulated signal is received, an audio voltage also will exist between point X and ground. For this reason, the automatic-volume-control voltage is obtained from point X as follows: A resistor of about 2 megohms and a capacitor of about 0.05 microfarad are connected in series between point X of Fig. 14 and ground. The resistor, which is between point X and the capacitor, limits the capacitor charging and discharging current. Thus, the capacitor voltage does not vary at the audio rate, but only in accordance with the relatively slow changes caused by variations in the signal transmission path; this capacitor voltage is used as the automatic-volume-control voltage.

Antennas for Amplitude-Modulation Radio Receivers. For point-to-point communication, the receiving antennas are engineered carefully (page 511). For standard amplitude-modulation broadcast reception (550 to 1600 kilocycles) the receiving antennas are usually of two types; first, "vertical straight-wire" types, and second, wire loops. The words were placed in quotation marks because an individual antenna may deviate far from this general classification.

Vertical Straight-Wire Receiving Antennas. The oncoming radio signal wave, which is vertically polarized in amplitude-modulation broadcast reception, induces a signal voltage in each elemental length of the antenna as shown in Fig. 15(a). These voltages add vectorially[13] and produce a current flow as indicated in (b). The current in the primary of the antenna coil induces a voltage in the secondary

which is the voltage impressed on the tuned-radio-frequency amplifier circuit previously discussed.

Loop Receiving Antennas. Many amplitude-modulation broadcast receivers utilize built-in loop antennas. There are many advantages

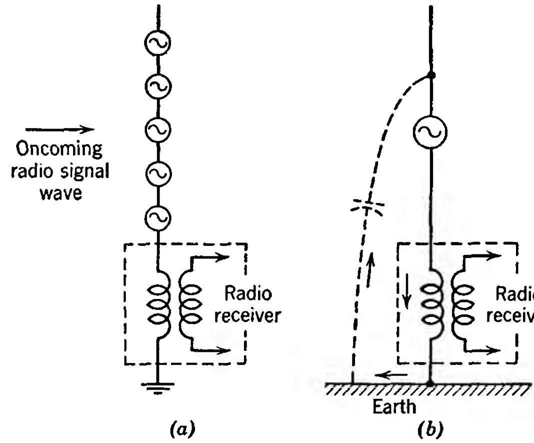


FIG. 15. The oncoming radio-frequency signal waves induce voltages in each unit length of antenna wire. The vector sum of these voltages is assumed to be concentrated at one point, and it forces an alternating signal current through the antenna coil of the radio-receiving set as shown by the arrows in (*b*). This current induces a signal voltage in the secondary of the antenna coil.

to this, including the possibilities of noise reduction (Chapter 14). Whether the loop receives signal because the vertical wires (or parts thereof) are parallel to the arriving electric lines of force or because the plane of the loop is perpendicular to the arriving magnetic lines of force is a common question, and one which is discussed in reference 14 (see also page 576). The signal voltage induced in the wire loop causes a current to flow around the loop which is tuned to series resonance with the capacitor (Fig. 16). The voltage drop across the capacitor is the signal voltage to be amplified (page 499).

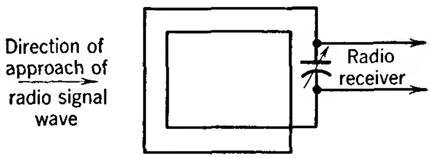


FIG. 16. Simplified circuit of a loop antenna. Many turns of wire are used at low radio frequencies.

On page 500 it was explained that a superheterodyne radio-receiving set contained a tuned-radio-frequency amplifier. In many super-

heterodynes this amplifier is omitted, and a tuned circuit or the tuned loop is used to give some initial selectivity. In very inexpensive (and often unsatisfactory) superheterodynes all tuning is omitted preceding the frequency converter. Such sets are bothered with the reception of **image frequencies** and with **cross modulation.** If the tuning ahead of the converter is omitted and if a 1,000,000-cycle station is being received, a station outside the broadcast band on an "image frequency" of 1,910,000 cycles will also be received. Cross modulation[12] is a form of interference caused by the modulation of the desired carrier by an undesired signal.

Amplitude-Modulation Radio-Telephone Systems. In this chapter and in preceding chapters the basic principles of amplitude-modulation and radio-telephone apparatus, such as antennas, transmitters, and receivers, have been discussed. In the pages immediately following radio-telephone systems that are of practical importance and use amplitude modulation will be considered.

Standard Amplitude-Modulation Broadcast Systems.[15] Standard amplitude-modulation broadcast systems operate in the frequency band of 550 to 1600 kilocycles. These systems are so common and have been discussed so extensively that little additional information need be added. There are (1949) approximately 2000 such broadcast stations in the United States.[16]

The microphones and studio facilities used were discussed in Chapters 2 and 4. The broadcast stations are often connected to network programs transmitted over open-wire lines and cables (Chapters 6 and 7). The circuits usually employed pass a band from about 100 to 5000 cycles, although network circuits are available that will provide transmission from approximately 50 to 8000 cycles. In the future most programs will be transmitted by telephone carrier systems.

Three classes of broadcast channels have been established;[15] they are **clear channels, regional channels,** and **local channels.** The classification of standard broadcast stations[15] is summarized as follows: **Class I stations** that are dominant stations and have from 10 to 50 kilowatts output; **Class II stations** that are secondary stations and radiate from 0.25 to 50 kilowatts; **Class III stations** that are regional stations and have power outputs from 0.5 to 5 kilowatts; and **Class IV stations** that are local stations and operate from 0.1 to 0.5 kilowatt. Each of the stations enumerated has other important distinguishing characteristics.[15] The types of services rendered by broadcast stations have been defined[15] as **primary service,** which is characterized by steady ground-wave propagation and has negligible fading and interference from other stations (page 572), **sec-**

ondary service, characterized both by ground-wave and sky-wave reception and by fading and some interference from other stations; and **intermittent service,** which is rendered by the ground wave and extends from the limit of the primary service area outward until it has no service value.[15]

Transoceanic Low-Frequency Radio-Telephone Systems. The first radio-telephone link for connecting land telephone systems was established in 1920 between the coast of California and Santa Catalina Island. The first transoceanic radio-telephone system was installed between New York and London in 1927. About 500,000 overseas

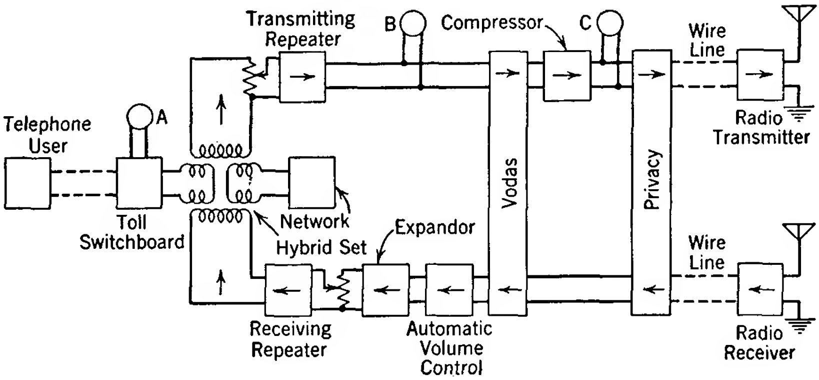


FIG. 17. Terminal arrangement for the long-wave transatlantic radio-telephone system. Instruments are shown at *A*, *B*, and *C* for monitoring purposes. (Reference 17.)

radio-telephone calls are now made annually. A description of the original radio-telephone amplitude-modulation system will be given because it embodied many novel features. The system to be described has for years been a standby for the short-wave systems and is used when traffic is heavy or during adverse ionospheric conditions when short-wave stations do not operate satisfactorily.

A schematic of the low-frequency, long-wave (60,000 cycles, 5000 meters) transatlantic radio-telephone circuit is shown in Fig. 17, and its operation[17] is as follows: The audio-frequency telephone signals produced by the speaker's voice pass over the connecting telephone line, through the toll switchboard, and into the hybrid coil. The signal voltage is impressed on the transmitting repeater where it is amplified, and then it passes through the Vodas, the compressor, the privacy equipment, and then over a wire line to the radio transmitting equipment. Since the radio transmitting and receiving equipment may be

located some distance apart to prevent interference, the privacy equipment, the Vodas, and the other associated apparatus are terminal equipment located at one point.

The **Vodas**[18] (voice-operated device, anti-singing) is a device which is automatically operated by the voice waves and which in effect short-circuits the *receiving* circuit at the *sending station* during speaking and then immediately returns it to the normal condition for receiving the answer from the distant station. In this way, feedback from the transmitting antenna of one channel to the receiving antenna of the same terminal is prevented, and two-way conversation is possible over the same band of radio frequencies.

The **compressor** is a vacuum-tube circuit for compressing the volume range of the speech signals at the sending end before transmission by radio to the distant station. At the receiving station the signal is passed through an **expander** which returns the signal to its original relations. These two units comprise the **Compandor,** which reduces the interference due to static.[17] It is apparent that, *if* the voice could be transmitted without *any* variations in signal volume, then the signal level could be maintained (most of the time) far above the static level. During portions of the speech where the signal is weak, static interference is bad because the signal-to-noise ratio is low. Thus, if it is possible to compress the speech and transmit it in this compressed form to the distant station where it is expanded to its original relations, less interference from static will result.

The **privacy equipment** "scrambles" the speech, rendering it unintelligible unless special "unscrambling" equipment is used. Several such systems are possible, among which is a **frequency inverting method.** Thus if a 50,000-cycle carrier is modulated by a 3000-cycle wave and the lower sideband only is selected with a filter, an output of 47,000 cycles will result. If this wave is now modulated by a 42,000-cycle wave and the lower sideband again selected with a filter, a 5000-cycle wave will result. If the original modulating wave had been 200 cycles, then the output of the second filter would be a 7800-cycle impulse. Thus, if the voice-frequency band used in commercial telephony varies from 200 to 3000 cycles and if this complex wave were impressed on the filter, the lower frequencies would appear as high-frequency components, and vice versa, in the output of this scrambling or inverting process. Also, the frequencies can be shifted from time to time to render "decoding" more difficult.

Single-sideband transmission is used in the transoceanic long-wave system. This requires a smaller frequency band and saves power and power-handling capacity. The power radiated in the sideband is

about 200 kilowatts, or some 100,000,000 times that in the original speech signal.[19, 20] The so-called multiple-tuned antenna is used for transmission, and the wave antenna (page 479) is used for re-

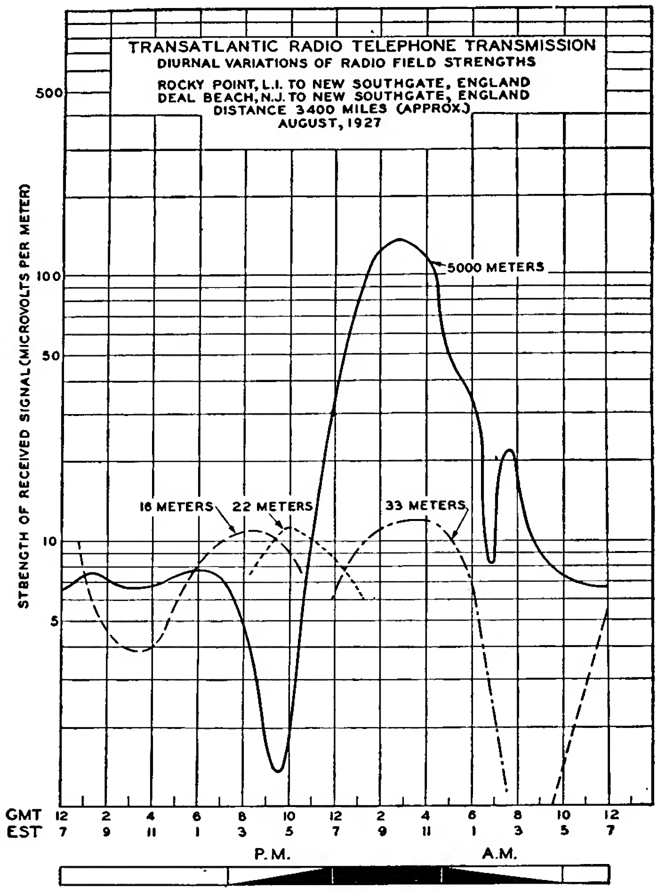


FIG. 18. Long-wave curve and corresponding short-wave curves of received field strength. (Courtesy Bell Telephone System.)

ception. Typical characteristics of the transatlantic path are shown in Fig. 18.

Transoceanic High-Frequency Radio-Telephone Systems. The overseas radio-telephone connections of the Bell System are shown in Fig. 19. With the exception of the long-wave system just described, the facilities of Fig. 19 are provided over short-wave amplitude-modulation systems.[21] Because of the large number of systems in use and

because of changes made from time to time, it is not feasible to attempt to describe the systems in detail. Instead, an overall picture of short-wave radio-telephone facilities will be presented.

The terminal arrangement of the equipment in a typical system is similar to that shown in Fig. 17. In the early systems Compandors were not used, noise reduction being achieved by voice control at the receiver.[22] This is based on the principle that the presence of background noise reduces the sensitivity of the ear (page 41) and reduces the intelligibility of a signal. Loss is introduced into the circuit during the intervals of no speech. Then, when speech does occur, the ear will be more sensitive than if the noise had not been reduced, and the signal, even with the background of static, will be more intelligible.

The original short-wave transoceanic radio-telephone systems transmitted the carrier and both sidebands; this is a wasteful process. For equal peak amplitudes in the transmitter, a system transmitting only one sideband gives a theoretical improvement of 9 db in received signal-to-noise ratio over a station transmitting the carrier and both sidebands.[23] Notwithstanding the advantages of single-sideband transmission, it was not used in the early systems for reasons which will now be given.

For single-sideband transmission, the carrier must be supplied for demodulation at the receiver by a local oscillator or by some other stable local source. For satisfactory speech demodulation, the local frequency must be within ± 20 cycles of the correct value.[23] This offered no serious problem in the long-wave channel operating at 60,000 cycles, but such high stability was difficult to achieve in short-wave oscillators operating at several million cycles.

To avoid this difficulty in the single-sideband short-wave systems installed in 1935 for transoceanic telephone purposes, a **pilot frequency** was transmitted over the system, in addition to the single sideband. For this purpose, some of the carrier (greatly reduced) of the transmitting set was transmitted along with one of the sidebands to the distant receiver, and thus the carrier itself acted as the pilot frequency. This weak received carrier signal may be used to synchronize automatically the locally generated carrier at the receiving station, or it may be amplified at the receiving station and then combined with the received sideband in the demodulator. Further details of the transmitters and receivers are given in references 24 and 25.

Transmitting Antennas for Radio-Telephone Systems. Originally, resonant broadside arrays were used in these systems.[26, 27] These were elaborate and costly and operated at one frequency only. Because each transoceanic telephone system is provided with several

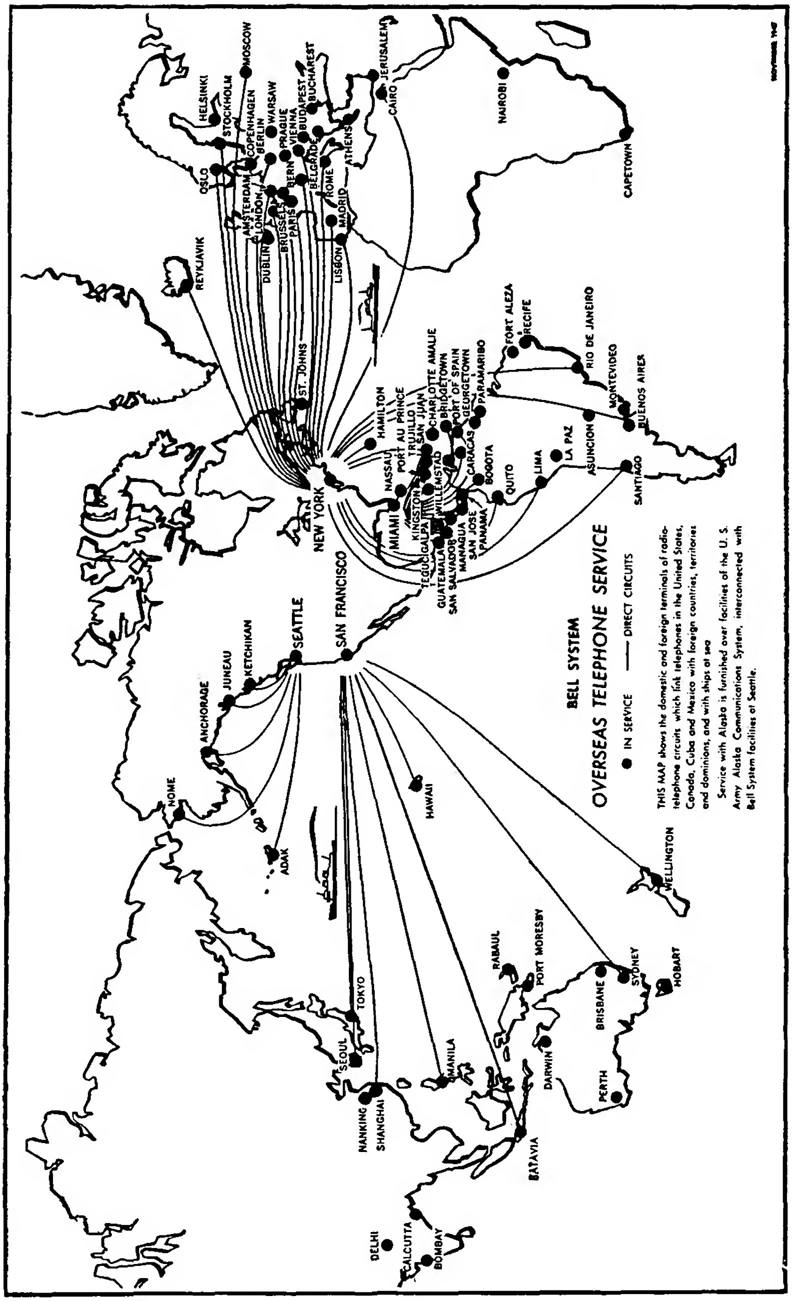


FIG. 19. Overseas radio-telephone circuits of the Bell System in use about 1949. (Courtesy Bell Telephone System.)

different frequencies for operation at different times (Fig. 18), it was necessary to have an antenna for each frequency. These resonant antenna arrays were soon replaced with non-resonant rhombics for both transmitting and receiving. For transmitting, a **twin rhombic** such as is shown in Fig. 20 is used. This gives characteristics better

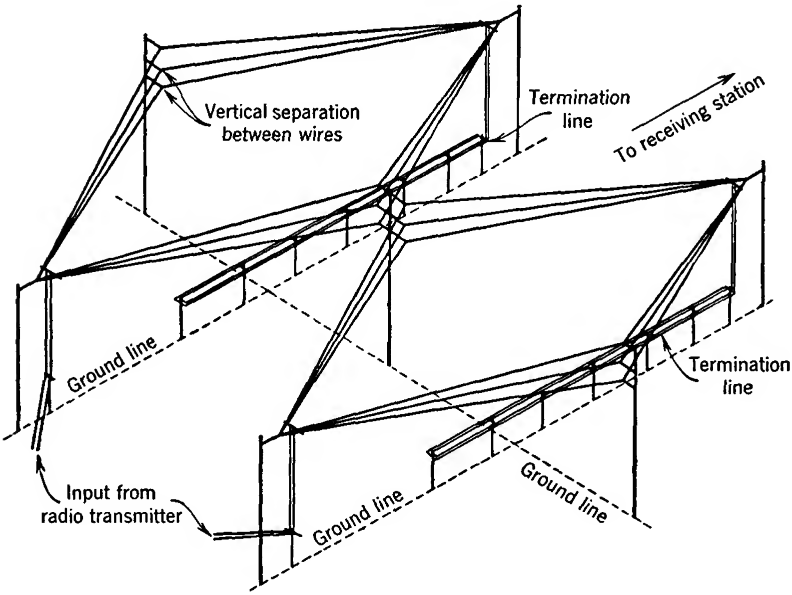


Fig. 20. A twin rhombic transmitting antenna is used in transoceanic radio-telephone circuits. This antenna produces highly directional transmission in the 5 to 25 megacycle band. (Courtesy American Telephone and Telegraph Co.)

than a single large rhombic. On page 482, the rhombic antenna was shown terminated with a resistor to prevent wave reflection. The "termination line" of iron wire shown in Fig. 20 serves the same purpose and is used with large transmitting antennas.

Receiving Antennas for Radio-Telephone Systems. The rhombic antenna (page 482) is most receptive to a wave coming in at a small angle with the horizontal. This is important for at least two reasons: *First,* the antenna is less sensitive to sources of noise (such as automobile ignition systems) on the surface of the earth; *second,* by altering the antenna dimensions this angle of maximum sensitivity can be changed. For maximum reception, the angle at which the antenna is most sensitive should coincide with the angle at which most of the electromagnetic energy arrives at the receiving antenna location from

the distant transmitting station. Thus by altering the dimensions the rhombic antenna may be *steered*[28] to the most favorable direction. By this action, fading is reduced. Rhombic antennas, which remain mechanically fixed in a horizontal plane, may also be steered by electrical changes in the receiving circuit. Thus, as many as sixteen rhombic antennas are combined into an array called a Multiple Unit Steerable Antenna or MUSA, and such an array gives decided advantages in reception over a single rhombic antenna unit for the following reasons.[21, 29]

Noise due to static and selective fading due to phase interference of signal components arriving over different paths are factors seriously affecting radio reception. The signal-to-noise ratio can be increased by using receiving antennas of sharper directional characteristics, but such antennas may be so highly directional as to "miss" the signal as it fluctuates, and they may thus have bad selective fading.

Studies have shown that what is regarded as a received radio signal is in reality the combination of several signals arriving at the receiving antenna over various transmission paths. The ordinary directional receiving antenna is "blunt" enough to receive substantially all these components. Since they are of varying magnitudes and phases, however, their combined signal is not constant but fades, thereby causing variations and accompanying distortion in the receiver output.

With the MUSA system, the individual signals picked up by a number of rhombic antennas are properly phased and then combined to give a large signal voltage. Strong incoming signals are separately "aimed at" by the individual MUSA units, and the received components are properly phased and then combined. Decided improvements, resulting in increased reliability of the short-wave systems, are obtained with MUSA arrays.

It has been found that with high-frequency radio systems the received signal at a *given frequency* may momentarily grow weak at a certain location but may actually grow stronger at a point only a few hundred feet away. Also, at a *given location* the signal received on one frequency may momentarily grow weak, but the signal strength at the same location may become greater for a frequency a few hundred cycles different. **Diversity receiving systems** have been developed for both radio-telephone and radio-telegraph systems to take advantage of these facts.[30, 31]

Systems are in use in which the signals from the same transmitting station, but received by two or more antennas spaced some distance apart, are properly combined to operate one receiving unit. These are termed **space diversity receiving systems.** Advantage is taken of the

fact that transmission may be poor on one frequency but excellent on one a few hundred cycles different in the **frequency diversity systems,** in which the same receiving equipment is used but the signal is transmitted simultaneously at slightly different frequencies.

Ship Radio-Telephone Systems. Experimental work, directed toward the establishment of ship-to-shore two-way telephone communication, by means of which a person on a ship at sea could be connected by radio with the land telephone systems of the United States, was started in 1919. Although these studies indicated the technical feasibility of the plan, it was dropped in 1922. Experiments were resumed some years later, and beginning in 1929 two-way telephone service was established with several of the large transatlantic liners.[32]

Two general types of systems now exist: those for providing radio-telephone connections to transoceanic ships[33] and those for providing coastal and harbor service.[34] The systems for service with transoceanic ships are similar to the short-wave systems described previously and employ amplitude-modulation equipment. The coastal and harbor equipment also is amplitude modulated, but it is less elaborate than the other systems.

Miscellaneous Radio-Telephone Systems. In addition to the services just described, amplitude-modulation radio-telephone systems are used for private communication networks such as those used by police, aircraft operating companies, and others. Such equipment is also used in emergencies for bridging gaps in wire telephone lines and for providing telephone toll services across bodies of water.[35] Several miscellaneous radio services, to be considered later, now use angle modulation.

Amplitude-Modulation Radio-Telegraph Systems. These systems are used for maritime mobile, or ship, telegraph service and for point-to-point service over both land and sea. Amplitude modulation is used in a large number of the radio-telegraph systems, but there is a growing tendency to use other methods.

With amplitude-modulated systems, either **continuous** (or **undamped) waves,** or **interrupted-continuous waves,** are employed. With the first method, "spurts" of radio-frequency waves are sent out from the transmitter. With the second method, the "spurts" of radio frequency are interrupted, for example, 1000 times a second.

The transmitters[36] are similar in many respects to amplitude-modulated radio-telephone transmitters, spark transmitters no longer being used. Methods are provided for **keying** the transmitter, that is, turning it on and off in accordance with the signals to be sent. The

keying may be by hand, by printing-telegraph teletypewriters, or by tape transmitters. Amplitude-modulation telegraphy, which is being discussed now, is often called **on-off modulation,** or **on-off keying.**

There is an important trend[37, 38] toward perforated tape methods of sending, receiving, and switching, comparable to that used with wire telegraphy (page 337). The extent of this is indicated by Fig. 21.

The radio-telegraph receivers usually are of the superheterodyne, or double-detection, type but are provided with a second oscillator (often called a **beating oscillator,** or a **beat-frequency oscillator**) that will introduce a wave into the intermediate-frequency amplifier so that continuous-wave messages can be received. These receivers are generally familiar, examples being in most radio laboratories, and are called **communication receivers.**

Frequency and Phase Modulation. Early in the history of radio a search was made for methods of modulation that would require less space in the radio-frequency spectrum for transmission of the signal. For some time it was thought that, if the frequency of the carrier wave (instead of the amplitude) were varied by the modulating wave and if the deviation from the center frequency were made small, then communication would be possible over a narrower frequency band than with amplitude-modulated waves. The error in this reasoning was pointed out[39] by Carson in 1922.

Several theoretical papers appeared on frequency modulation, notably in 1930 by van der Pohl,[40] and in 1931 by Roder.[41] At about this time Armstrong did experimental work with frequency modulation. As a result, he was able to show that much reduction in received noise was possible if a rather large frequency deviation was used. His important 1936 paper[42] on the subject initiated the practical development of frequency-modulation methods. In 1940 Everitt summarized existing knowledge and techniques.[43] The literature now contains much information on the subject.

Angle Modulation. Angle modulation includes both frequency and phase modulation, which are closely related. **Angle modulation** is defined[2] as "modulation in which the angle of a sine-wave carrier is the characteristic subject to variation." Much of the information that follows is summarized from the papers by Roder[41] and Everitt.[43]

The fact has been stressed that modulation is merely a process by which the message or program to be transmitted is translated, or moved, to a band in the radio-frequency spectrum. Also, it has been pointed out that the so-called carrier wave in no sense "carries" the signal to the distant station.

The definition[2] of **phase modulation** or **PM** is "angle modulation in

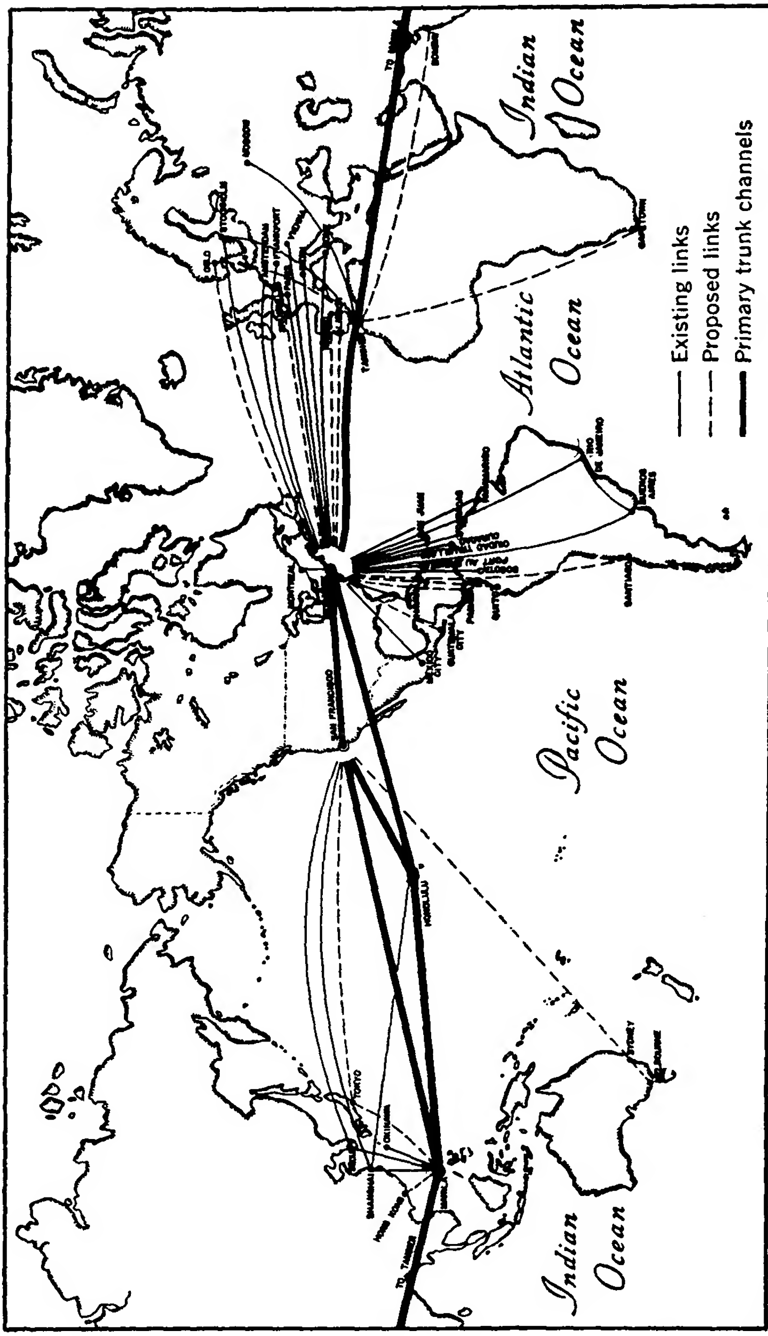


FIG. 21. International radio-telegraph tape-printing relay network. (Reference 38.)

which the angle of a sine-wave carrier is caused to depart from the carrier angle by an amount proportional to the instantaneous value of the modulating wave." The definition[2] of **frequency modulation** or **FM** is "angle modulation in which the instantaneous frequency of a sine-wave carrier is caused to depart from the carrier frequency by an amount proportional to the instantaneous value of the modulating wave."

To investigate phase and frequency modulation, assume that both a sinusoidal carrier wave and a sinusoidal modulating wave are impressed on a modulator that can cause amplitude, phase, or frequency modulation. The instantaneous value of the carrier voltage wave will be

$$e_c = E_{mc} \sin \theta_c = E_{mc} \sin \omega_c t, \tag{3}$$

and the instantaneous value of the voice-frequency modulating voltage wave will be

$$e_v = E_{mv} \sin \theta_v = E_{mv} \sin \omega_v t. \tag{4}$$

As has been explained, in amplitude modulation E_{mc} is caused to vary in accordance with the modulating frequency, and two sidebands, each containing the information to be transmitted, are created. A **phase-modulated sinusoidal wave**[2] is one "in which the argument contains a term whose wave form is similar to that of the signal to be transmitted." A **frequency-modulated sinusoidal wave** is one "in which the expression for instantaneous frequency contains a term whose wave form is similar to that of the signal to be transmitted." "The **instantaneous frequency**[2] of a sinusoidal wave whose argument is a function of time equals $1/(2\pi)$ times the derivative of the argument with respect to time."

Because in equation 3 $\omega_c = 2\pi f_c$, where f_c is the assigned carrier frequency and is a fixed value, this equation must be written in a more general form to show how phase and frequency modulation are accomplished. Thus,

$$e_c = E_{mc} \sin \theta'_c = E_{mc} \sin(\omega_c t + \phi). \tag{5}$$

In *phase modulation*, the angle ϕ is made to vary in accordance with the modulating signal. If m_p is the **phase deviation,** defined[2] as "the peak difference between the instantaneous angle of the modulated wave and the angle of the carrier," and is the maximum phase shift in radians produced by the sinusoidal modulating signal of equation 4, then the instantaneous value of a phase-modulated wave is

$$e_p = E_{mc} \sin(\omega_c t + m_p \sin \omega_v t). \tag{6}$$

In *frequency modulation,* the instantaneous frequency of the output wave of the modulator must be caused by the modulating signal to vary about the average frequency $f_c = \omega_c/2\pi$. The instantaneous frequency is by definition[2]

$$f_{\text{inst}} = \frac{1}{2\pi}\frac{d\theta'_c}{dt}. \tag{7}$$

Thus, from equation 5,

$$f_{\text{inst}} = \frac{1}{2\pi}\frac{d\theta'_c}{dt} = \frac{\omega_c}{2\pi} + \frac{1}{2\pi}\frac{d\phi}{dt}. \tag{8}$$

Because ω_c is constant, the modulating wave must vary $d\phi/dt$, and hence the instantaneous frequency is

$$f_{\text{inst}} = \frac{\omega_c}{2\pi} + m_f \sin \omega_v t, \tag{9}$$

where m_f is the **frequency deviation** and is[2] "the peak difference between the instantaneous frequency of the modulated wave and the carrier frequency." It has been called the frequency-modulation factor. From equations 8 and 9,

$$\frac{d\phi}{dt} = 2\pi m_f \sin \omega_v t, \tag{10}$$

and integrating,

$$\phi = -\frac{m_f}{f_v}\cos \omega_v t. \tag{11}$$

When this value is substituted in equation 5, the equation for a frequency-modulated wave becomes

$$e_f = E_{mc}\sin\left(\omega_c t - \frac{m_f}{f_v}\cos \omega_v t\right). \tag{12}$$

Comparison of Phase and Frequency Modulation. It will be noted that equation 6 for a phase-modulated wave is very much like equation 12 for a frequency-modulated wave, the important difference being that in phase modulation the maximum phase shift during modulation is given by m_p and is independent of the modulating frequency, but in frequency modulation the maximum phase shift is given by what has been called the **modulation index,**[2, 40, 41] m_f/f_v, which varies inversely as the modulating frequency f_v. The ratio "of the maximum frequency deviation to the maximum modulating frequency of the system" is now called[2] the **deviation ratio.** In phase modulation, m_p

is the maximum shift in phase that the audio signal of maximum intensity will cause, and, in frequency modulation, m_f is the maximum frequency shift that the audio signal of maximum intensity will cause. Accordingly, Roder stated[41] that "frequency modulation corresponds to phase modulation where the *amplitude* of the audio signal is inversely proportional to frequency of that signal." A system of phase modulation will produce frequency modulation, or vice versa, if the modulating signal voltage is thus predistorted. Also, a system for demodulating frequency-modulated waves will demodulate phase-modulated waves, and vice versa, if after demodulation the wave is

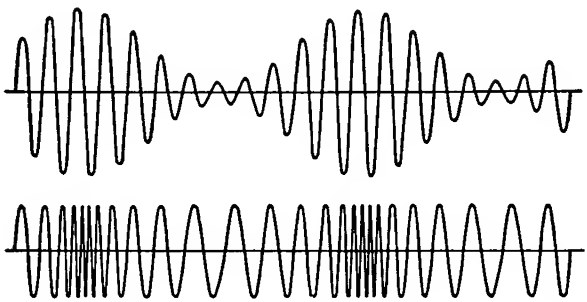


FIG. 22. An amplitude-modulated wave is shown above, and a frequency-modulated wave below. Each is modulated by a pure sine wave.

distorted. Another point brought out by Roder was that at corresponding intervals in amplitude modulation the carrier and sidebands are in phase, but in phase and frequency modulation, for small values of m_p and m_f/f_v, the carrier and sidebands are 90° out of phase.

The shapes of an amplitude-modulated wave and a phase or frequency-modulated wave are shown in Fig. 22. In phase modulation, the magnitude of the phase shift depends on the magnitude of the modulating signal, and the number of phase shifts per second depends on the frequency of the modulating signal. In frequency modulation, the magnitude of the frequency shift depends on the magnitude of the modulating signal, and the number of frequency shifts per second depends on the frequency of the modulating signal.

Sidebands were found to exist in an amplitude-modulated wave, and they also exist in a phase-modulated wave or frequency-modulated wave, *but in greater number;* in fact, in these two systems *an infinite number of sidebands exist* (at least theoretically). This is shown[41, 43] by expanding equations 6 and 12 in accordance with the theory of Bessel's functions. Figure 23 has been prepared[43] in this manner.

It shows that for a frequency-modulated wave, because of ratio m_f/f_v previously mentioned, a high audio-frequency modulating signal does not cause so large a *number* of appreciable sidebands (or side-frequencies for a sinusoidal modulating wave) as is caused by a modulating wave of lower frequency. It should be noted, however, that the overall bandwidth required in each instance is about the same and that it is wider than the maximum deviation in frequency for the high-frequency components of the audio-frequency modulating signal. The practical significance is this: Modern frequency-modulation broadcast stations are adjusted for a maximum frequency deviation, or maximum change in frequency with the strongest modulating

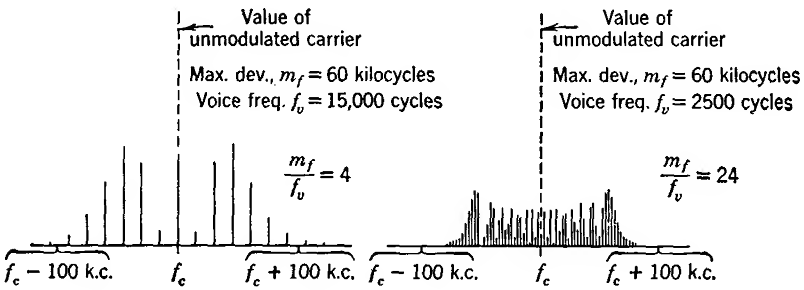


FIG. 23. These figures show that frequency-modulated waves contain a large number of sidebands. Note that in frequency modulation the magnitude of the component of carrier frequency, f_c, varies. (Adapted from reference 43.)

signal, of ±75,000 cycles around the assigned carrier frequency which is for broadcast in the vicinity of 100,000,000 cycles. But the maximum bandwidth required to pass the sidebands of appreciable magnitude is about 200,000 cycles wide. Thus, each frequency-modulation broadcast station is assigned a band in the frequency spectrum that is 200,000 cycles wide, and that is one important reason why frequency-modulation broadcasting is placed in the 100-megacycle region.

It is of interest to compare the bandwidth required for a phase-modulated wave and for a frequency-modulated wave. As Fig. 23 shows, for frequency modulation the bandwidth is about the same for any ratio of m_f/f_v. For comparable conditions, this is *not* true for a phase-modulated signal; for, as equation 6 indicates, the shift in phase is *not* inversely proportional to the modulating frequency. Thus, a phase-modulated signal will cover a much wider frequency band than a frequency-modulated signal will under comparable conditions. Frequency modulation is extensively used in practice, and

phase modulation is not. However, the two methods are so closely related that many of the frequency-modulated waves are produced by phase-modulation methods.

Methods of Producing Frequency-Modulated Waves. Many methods of frequency modulation are used in broadcast and other radio applications.

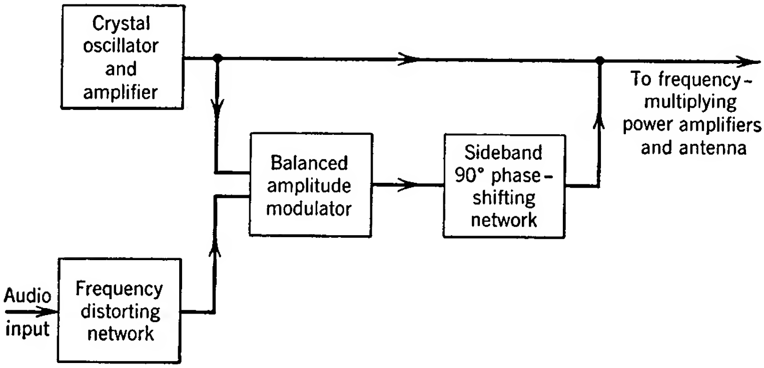


Fig. 24. Arrangement for synthesizing a frequency-modulated wave.

Indirect Frequency Modulation by Wave Synthesis. This method follows directly from Roder's investigation.[41] It was the system used by Armstrong,[42] and is used in modified form in broadcast trans-

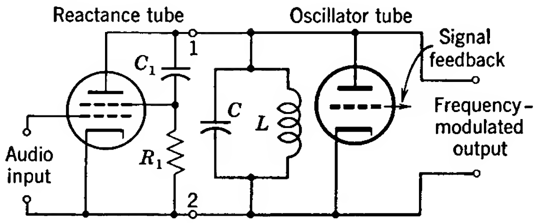


Fig. 25. Simplified circuit of a reactance-tube frequency modulator.

mitters. The block diagram of Fig. 24 shows an arrangement by which a frequency-modulated wave is synthesized, following statements given in the preceding section.

Direct Frequency Modulation Using the Reactance Tube. The reactance-tube method is shown in Fig. 25 and is the outgrowth of methods of controlling the frequency of the local oscillator in an amplitude-modulation superheterodyne receiver.[44, 45] It is defined[2] as **reactance modulation.** Briefly, the alternating voltage from the

oscillator existing across C–L causes a current to flow down through C_1, which is of small capacitance (and high reactance), and through R_1, which is of small resistance. This current and the voltage drop across R_1 will lead the voltage across points 1–2 by essentially 90°. Because in the circuit used the plate current is in phase with the grid voltage, the alternating plate current of the reactance tube that flows to the right of points 1–2 will lead the voltage across points 1–2 by essentially 90°. The magnitude of this current is varied by the modulating audio-frequency signal voltage connected as indicated. This is equivalent to connecting a voice-controlled capacitor across points 1–2, and hence the incoming modulating signal causes the oscillator output to be frequency modulated. The positions of R_1 and C_1 may be interchanged if the proper values are selected. Also, tubes in push-pull are used.[46, 47]

Frequency Modulation by Special Tubes. Several special tubes have been developed for producing frequency-modulated waves, one being the **Phasitron.**[48] In this device, electrons are emitted from a central cathode and flow out in the form of an electron disk toward surrounding cylindrical anodes. Below the disk of electrons are special electric deflecting electrodes composed of wires connected successively to the phases of a three-phase radio-frequency voltage. This voltage is obtained from phase-splitting networks that are driven by a single-phase crystal-controlled oscillator. These deflecting wires will move the various parts of the electron disk up or down, depending on the polarities of the various wires. This will cause the edge of the electron disk to be "scalloped" sinusoidally and, furthermore, the scallops will move around the edge of the disk at a definite rate because the three-phase radio-frequency voltage impressed on the wire electrodes will produce a "revolving" electric field.

Two cylindrical anodes are provided. The innermost has square holes in it, and the two anodes are connected to a tuned output circuit. As the electron disk rotates, electrons flow first to one anode and then to the other, causing a high-frequency current flow in the tuned output circuit. The modulation is effected by passing the modulating signal current through a coil surrounding the entire tube structure. The electron disk, with its sinusoidal scallops, is advanced in phase and retarded in phase, by the magnetic field produced by the modulating signal current in the surrounding coil, thus producing phase modulation. This current is predistorted so that the output wave will be frequency modulated.

Frequency Modulation by Non-Linear Coils. Among the miscellaneous methods of frequency modulation is a circuit using a non-linear

coil.[49] This system has been used extensively in systems of communication (as distinguished from broadcast) using frequency modulation. Actually, the circuit produces phase modulation, and predistortion is necessary.

The Demodulation of Frequency-Modulated Waves. Several systems have been developed. Two methods that are successfully employed will be discussed.

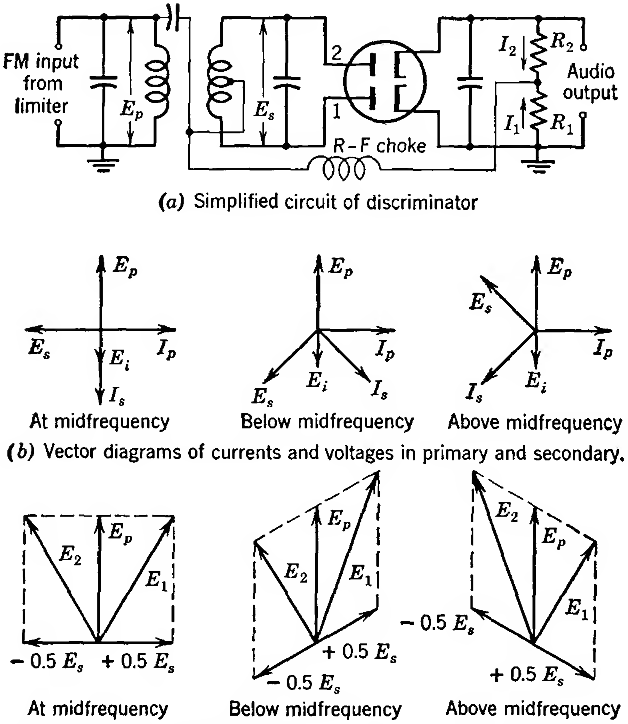
The Limiter and Discriminator.[2] The **limiter** is a device that "clips off" the received and amplified frequency-modulated signal so that no amplitude variations (such as might be caused by static) occur in the wave envelope. Several types of limiters are used, singly or in combination; among these are the crystal detector (page 272), circuits using vacuum tubes that limit because of grid-current flow through a resistor, and circuits using tuned amplifiers that saturate because of low plate voltage.[43] Because it removes unwanted amplitude variations, the limiter is very important.

The **discriminator** is the demodulating device that is actuated by the frequency-modulated radio signal and produces a replica of the original modulating signal. The discriminator is an outgrowth of frequency-control methods.[44, 45] Modifications of the discriminator circuit of Fig. 26 are used.

The operation of the discriminator *when the received signal is at midfrequency* is as follows: After passing through the limiter, the radio-frequency signal that is frequency modulated is impressed as indicated in Fig. 26. The primary is tuned to the midfrequency, and, if at the instant under consideration the frequency is at midfrequency, the vector diagram at the left in (*b*) applies. The impressed voltage E_p causes the 90° lagging current (approximately) to flow in the primary coil. This current induces a 90° lagging voltage E_i *in series* in the secondary coil that also is tuned to the midfrequency. The secondary current I_s will be in phase with this voltage when the signal is at midfrequency. A voltage drop will be caused across the secondary capacitor, and this will lag the secondary current by 90°. The connections are assumed to be such that the voltage between the anode and the cathode of each tube is the vector sum of the primary voltage E_p and one-half of the secondary voltage E_s. The capacitor connecting the primary and the center of the secondary is sufficiently large to have negligible reactance. When the signal is at midfrequency E_1 and E_2 are equal, the rectified currents I_1 and I_2 are equal, and, since R_1 equals R_2, no resultant voltage will exist at the audio-output terminals.

The operation of the discriminator *when the received signal is below*

midfrequency is as follows: For this frequency the secondary will be out of resonance, capacitive reactance will predominate, I_s will lead E_i, and the secondary voltage E_s will be as indicated. This causes diode voltage E_1 to exceed diode voltage E_2, rectified current I_1 will exceed



(*a*) Simplified circuit of discriminator

(*b*) Vector diagrams of currents and voltages in primary and secondary.

(*c*) Showing the voltages E_1 and E_2 across the diode elements as the frequency-modulated signal frequency varies.

FIG. 26. Circuit and vector diagrams for a discriminator used to demodulate frequency-modulated signals.

I_2, and, since R_1 equals R_2, the lower audio-output terminal will be positive and the upper terminal will be negative.

The operation of the discriminator *when the received signal is above midfrequency* is as follows: The secondary will be out of resonance for this frequency and inductive reactance will predominate, causing the secondary current I_s to lag the voltage E_i. This results in diode voltage E_2 exceeding E_1; also, rectified current I_2 will exceed I_1, and the upper audio-output terminal will be positive and the lower negative.

At the frequency-modulation transmitter, one cycle of modulating

audio signal voltage causes a frequency shift below midfrequency and a shift above midfrequency. This will cause one cycle of audio-output discriminator voltage. At the transmitter a strong modulating voltage will cause a large frequency shift, and, at the receiver, a large frequency shift will result in a greater lead, or lag, and a stronger audio output.

In some circuits the capacitor connecting the primary and the secondary of Fig. 26 is omitted, and a coil is placed in the wire leading to the center of the secondary. This coil is coupled to the primary and serves to pick up a reference voltage that corresponds to E_p of Fig. 26(*b*) and (*c*).

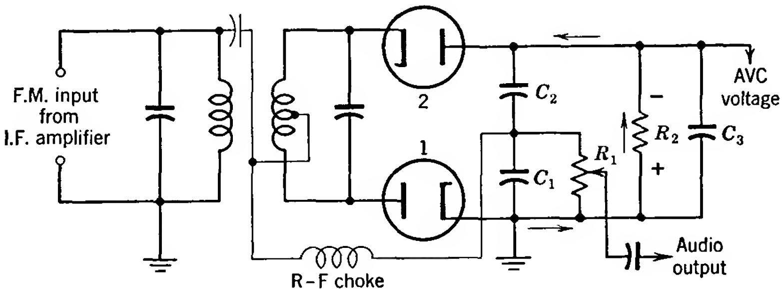


FIG. 27. Circuit of the ratio detector, a demodulator for frequency-modulated signals that need not be preceded by a limiter.

The discriminator is satisfactory, particularly if it is modified as just explained, but it must be preceded by a limiter. Because the limiter reduces the signal to the point at which all normal amplitude variations are eliminated, the intermediate-frequency amplifier preceding the limiter must have much gain; also, two limiter stages are sometimes used for more complete limiting action.

The Ratio Detector. The ratio-detector of Fig. 27 requires no limiter. Note that one diode is reversed and that other significant changes have been made from Fig. 26. Also, the capacitor between primary and secondary is usually omitted, and, as for some discriminators, another secondary coil is placed in series with the wire to the center of the secondary.[50] The vector diagrams of Fig. 26 apply to the ratio detector.

Capacitor C_3 of Fig. 27 is often about 8 microfarads, and the circuit is such that the voltage across C_3 can vary only slowly. Assume that a station is unmodulated and is, accordingly, on center frequency. If the station produces a strong signal (the limiter is omitted when a ratio detector is used), a large rectified current will flow as shown by

the arrows, and capacitor C_3 will charge slowly to a relatively large voltage value. If the station produces a weak signal, a small rectified current will flow and capacitor C_3 will discharge slowly to a low value. This makes available an **automatic volume-control voltage** between the terminal *AVC* and ground. Note that at the center frequency no current flows through the *R–F* choke or resistor R_1.

If the transmitter is frequency modulated and the signal falls below the center frequency, the radio-frequency voltage across diode 1 will increase and that across diode 2 will decrease (Fig. 27). This will cause the current in diode 2 to fall and that in diode 1 to rise; this causes flow through the *R–F* choke and through resistor R_1. The voltage across capacitor C_2 will fall, and that across C_1 will rise, but the sum of these voltages will equal that across C_3. When the frequency rises above the midfrequency, the action is reversed. Thus, an audio voltage, which corresponds to the frequency shifts in the received-modulated signal, exists across resistor R_1. If a *sudden* change occurs in the *magnitude* of the received frequency-modulated signal, the magnitude of the demodulated output voltage remains fixed because capacitor C_3 holds the voltage constant across capacitors C_1 and C_2. Only the ratio of these voltages can change. Thus a limiter need not be used.

Frequency-Modulation Radio-Telephone Transmitters. Transmitters for frequency-modulation are not standardized like those for amplitude modulation. They differ fundamentally in two important ways; first, the method of modulation; and second, the method of stabilizing the center frequency. Several methods of frequency modulation have already been discussed. The methods of frequency stabilization are too varied and detailed for consideration. The center frequency must not drift more than 2000 cycles from the assigned value.[51] A summary of the methods is given in the reference accompanying Fig. 28.

A block diagram of a frequency-modulation transmitter is shown in Fig. 28. The modulation is accomplished in most transmitters at a carrier frequency of several hundred thousand cycles and with a fairly low maximum frequency deviation. Also, modulation is accomplished at a low power level using receiving-type tubes; this is a decided advantage of frequency modulation. The frequency multipliers increase both the frequency and the frequency deviation to the correct values, and the power amplifier increases the power level to the proper amount. Because the frequency-modulated wave has constant amplitude, the class C power amplifier operates at constant load. At the very high frequency of approximately 100 megacycles, special

tubes and circuits are used. Input and output tuned circuits of the "coaxial-cable type" instead of inductors and capacitors are common.

Although this section has been devoted to frequency-modulation telephone transmitters, it should be mentioned that frequency-modu-

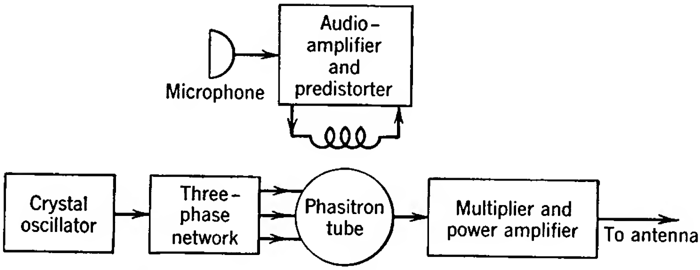


FIG. 28. Block diagram of a frequency-modulation radio-broadcast transmitter using the Phasitron tube. For block diagrams of other such transmitters see chart "Frequency Modulation Systems," supplement to *Electronic Industries*, April, 1946.

lation radio telegraph transmitters also are used extensively. These employ what is called[2] **frequency-shift keying** or **FSK** (see also page 335).

Frequency-Modulation Radio-Telephone Receivers. These are superheterodynes built to operate at the very high frequencies em-

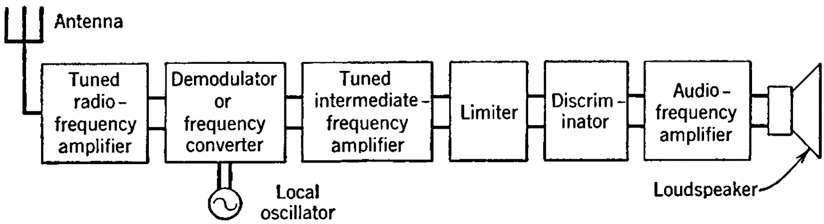


FIG. 29. Block diagram of the superheterodyne radio-receiving set for frequency-modulated signals.

ployed. The intermediate-frequency amplifier operates at about 10 megacycles in many sets. The set shown in Fig. 29 includes a limiter and a discriminator, but, of course, a ratio detector may be used instead.

In frequency-modulation broadcast transmitters **pre-emphasis circuits**[2, 51] are used which emphasize the high frequencies, and **de-emphasis circuits**[2] are used in the receiver to return the frequencies to their proper relative strengths. Such circuits, which are used to

further overcome noise, have been omitted from the drawings shown here. Frequency-modulation receivers[52] using what is called a **fremodyne circuit** for detection are also used. This circuit employs the principle of superregeneration discussed in many books on radio.[5] The **folded-dipole antenna**[5, 6, 53] is often used with frequency-modulation receivers.

Vehicle Radio-Telephone Systems. The first experiments in radio communication with automobiles were made in 1921 in connection with police systems.[54] Beginning in 1947, two-way radio-telephone service was given from the public telephone system to vehicles. A telephone user may place a call for a vehicle in much the same manner as a toll call is placed. One or more 250-watt transmitters cover a given area, and several receiving stations are located in the area.[55] At the beginning of 1949, about 5000 vehicles were equipped for this growing service.

Radio-Relay Systems. Since the beginning of electrical communication, efforts have been directed toward increasing the number of messages transmitted from point to point over one two-wire circuit. These efforts have resulted in the coaxial cable that will pass as many as 600 telephone messages simultaneously. With the development of radio, attention was given to the development of **point-to-point radio-relay systems** that would each pass a large number of messages. Coaxial cables and radio-relay systems are in economic competition. The United States is now (1949) spanned with a coaxial system, and installation of a transcontinental point-to-point radio relay system is in progress. These systems may be used either for telephone purposes or for television.

Radio-relay systems operate at superhigh frequencies, using wavelengths of approximately 10 centimeters. They use small amounts of power and highly directional antennas that are of the electromagnetic horn,[56] parabolic,[57] or lens[58] types. The antennas are mounted on tall buildings, towers, and hills.

Surprising as it may seem, such systems have been used for some time, the first probably being a two-way channel installed in 1931 across the straits of Dover and using Barkhausen, or positive-grid, oscillators, page 306, at approximately 1725 megacycles and a wavelength of 17.4 centimeters.[59]

Recent literature on point-to-point radio relay systems has been extensive.[60, 61, 62, 63, 64] In the paragraphs immediately following, the New York–Boston radio relay system will be described briefly.

Two basic types of equipment are required: the **terminal equipment,**[64] by means of which the messages or television programs to be

transmitted are translated to the superhigh frequency region, and the **micro-wave repeaters,**[62] which receive the complex signal after it has traveled about 30 miles, amplify the signal, and direct it on to the next relay repeater station.

The New York–Boston system operates at about 4000 megacycles (wavelength about 7.5 centimeters). Two two-way channels are pro-

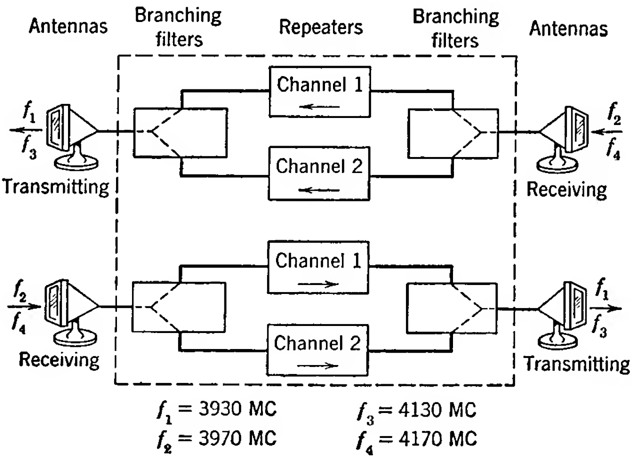


FIG. 30. Block diagram of a repeater of the New York–Boston radio-relay system. Frequencies are translated, or shifted, at repeater stations so that high-energy level and low-energy level signals will not exist at the same frequency at the same point. (Reference 63.)

vided, each capable of handling a television program extending from 30 cycles to 4.5 megacycles; or, the system can be used to transmit hundreds of telephone messages furnished by a type L carrier system (page 432). The average length of the repeater sections is 27.5 miles, 35 miles being the longest. At the input terminal equipment, the incoming signal is frequency modulated and emerges at a 65-megacycle center frequency. By modulation methods, this is raised to 3930 megacycles, is amplified, and passes through waveguides to the transmitting antenna, which is a metallic lens in the mouth of an electromagnetic horn. Parallel strips of metal of various dimensions act as waveguides and produce a focusing effect. At the receiving terminal the process is reversed. At the repeater stations, shown in block form in Fig. 30, the signal is received by an antenna as previously described, is reduced to an "intermediate frequency" of 65 megacycles, is amplified, then translated to about 4000 megacycles, is amplified in a "radio-

frequency" amplifier, and finally is retransmitted to the next station using the same type antenna. These translations are necessary because the velocity-modulated amplifying tubes (page 306) for low-level amplification at 4000 megacycles proved to be too noisy. Conventional tubes are used for low-level amplification at 65 megacycles, and velocity-modulated tubes (page 308) are used for high-level amplification at 4000 megacycles.[62] In the New York–Chicago link of the transcontinental radio-relay system, and in similar installations, special high-frequency triodes instead of velocity-modulated tubes are used.*

The system just described is operated by the Bell System. Another point-to-point microwave relay system is operated by Western Union over a triangle connecting New York, Pittsburgh, and Washington. Frequencies of about 4000 megacycles are used. Klystrons that are frequency modulated are used to drive small dipole antennas with parabolic reflectors. The Klystron is frequency modulated by a resultant signal composed of many telegraph messages. The power output of the Klystron is about 100 milliwatts. A detailed description of the system is given in *Western Union Technical Review*, July, 1948, Vol. 2, No. 3.

Pulse Modulation. Methods that are generally termed **pulse modulation** are under intensive experimental study and are used to some extent in practice, particularly with radio-relay systems. A study of this subject is complicated by the fact that it is in the developmental state. It is not evident which systems will eventually achieve wide usage, although the last one to be described is very promising. When pulses instead of a sine-wave carrier are used, a large number of possible modulation methods are available.

Perhaps what is called pulse modulation could be described as pulse sampling, and the general method described as communication by samples.[65] Several possible pulse-modulation schemes will be described.[65, 66, 67]

In Fig. 31 is shown a chart for explaining certain types of pulse modulation. Sinusoidal modulating signals are shown for convenience. In **pulse-amplitude modulation** or **PAM**[2] the signal to be transmitted is sampled at regular intervals, and pulses of equal width but of varying amplitude are transmitted. The receiver must measure the amplitude, or voltage, of each pulse and then reproduce the modulating signal voltage.[65]

In **pulse-duration modulation,** or **PDM,**[2] also called pulse-length modulation, or PLM,[65] and pulse-width modulation, or PWM,[66] the

* *Electronics,* April, 1949, p. 170.

signal to be transmitted is sampled periodically, and pulses whose duration, or length, or width, which vary with the instantaneous magnitude of the signal, are transmitted. The receiver must measure the time duration of the arriving pulses and reconstruct the original modulating signal from these time measurements.[65]

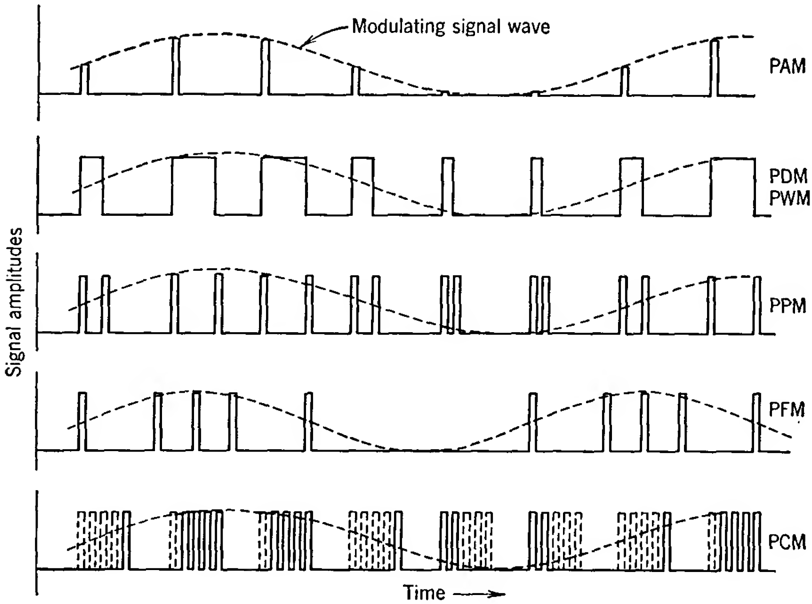


FIG. 31. Approximate shapes and arrangements of signals in pulse modulation with the sinusoidal waves shown as dotted curves. Height of pulses does not, in general, equal height of modulating wave. Also, many more pulses per cycle may be transmitted. These diagrams are to convey a general picture of the processes and should not be regarded as oscillograms. (Based on a diagram appearing in Reference 66.)

In **pulse-position modulation,** or **PPM,**[2] the amplitude of the pulses and their width remain constant, but the time at which they are transmitted depends on the instantaneous magnitude of the signal. The receiver must measure the time arrival of the signal and reconstruct the original modulating signal from this information.[65]

In **pulse-frequency modulation,** or **PFM,** the repetition frequency of the pulse is varied.[66]

In **pulse-code modulation,** or **PCM,**[65, 68, 68(a)] also called pulse-count modulation,[66, 67] the amplitude range of the modulating signal is divided into a number of discrete levels, and these are sampled at

regular intervals. Pulses bearing the information as to signal level are transmitted in groups, one group for each sample, and one or more pulses of a group are omitted to deliver specific data regarding the amplitude range of a sample. This is at present the most important system. The receiver must note the presence or absence of individual pulses in a group and reconstruct the modulating signal.

The circuits used for these systems of communication are different from ordinary circuits previously discussed, and, in fact, some of them involve the use of special vacuum tubes.[69, 70, 71]

The preceding paragraphs are *descriptions*. Definitions[2] applying to certain types of pulse modulation have been formulated as follows:

Pulse-Amplitude Modulation or PAM. Modulation in which the modulating wave is caused to amplitude-modulate a pulse carrier.

Pulse-Time Modulation or PTM. Modulation in which the values of instantaneous samples of the modulating wave are caused to modulate the time of occurrence of some characteristic of a pulse carrier.

Pulse-Duration Modulation or PDM. Pulse-time modulation in which the value of each instantaneous sample of the modulating wave is caused to modulate the duration of a pulse.

Pulse-Position Modulation or PPM. Pulse-time modulation in which the value of each instantaneous sample of a modulating wave is caused to modulate the position in time of a pulse.

Facsimile Transmission. Many systems have been devised for electrically transmitting drawings, figures, photographs, or other information over wire lines and cables, or by radio. A system of facsimile for transmission over wire circuits[72] was placed in commercial operation over telephone lines in 1925. This was replaced in 1935 by an improved method[73] which involved modulation by optical means. Picture transmission for news purposes by both wire and radio methods now is commonplace.

The principle involved in many systems of facsimile transmission is to revolve slowly a cylinder on which the picture or other information to be transmitted is attached. In some systems a light beam and phototube are used to convert the "lights" and "shades" of the image into electric signals. In other systems, a special paper (page 339) is employed. At the receiving station a synchronized reverse process is used to record the information by recreating, on paper, an image of the material scanned.

Television. Experimental work on television has been in progress for many years. A *two-way* television system[74] that transmitted over wire telephone circuits was operated for some months about 1930, and at approximately the same time television in colors was also

demonstrated.[75] During the years following these experiments, television has been the subject of much research, and several successful methods have been developed. Television shows great promise, and there are some who predict that television is destined to capture the present radio-broadcast audience, particularly in city and suburban areas. This remains for the future to disclose.

The Image Orthicon. Much of the success of television is due to the development of the **Image Orthicon**[76] by the Radio Corporation of America.

A simplified schematic diagram of the Image Orthicon is shown in Fig. 32. The light reflected by the object or scene to be televised is

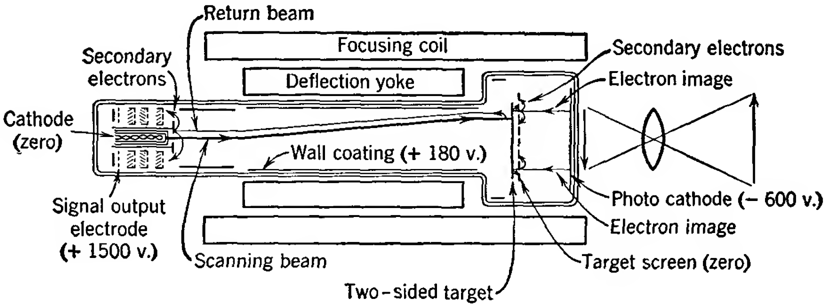


FIG. 32. Diagram of the R.C.A. Image Orthicon, a television camera tube used extensively. (Courtesy American Institute of Electrical Engineers.)

focused on the photocathode, which throws off electrons, forming an "electron image" of the object or scene. The two-sided target and target screen (Fig. 32) are at +600 volts *with respect to the photocathode,* and the electron image travels to the target. The electrons constituting the electron image are maintained in their respective places by the focusing coil that holds them in parallel rays.

The two-sided target of Fig. 32 is of low-resistance semiconducting glass. It is so thin that electric charges flow readily through the glass but do not flow "sideways" with ease. The electrons in the electron image striking the glass target knock out secondary electrons, and a positive charge pattern is formed on one side of the glass. This pattern is an electric-charge image of the visible image to be televised.

A beam of electrons from an electron gun is directed at the other side of the target, and this beam is caused to move in lines back and forth at various levels (somewhat as the page of a book is read), thus **scanning** the side of the two-sided target nearest the electron gun.

Because the glass conducts from one side to the other, if secondary emission from the "front" side maintains a given area on this side positive, the corresponding area on the "back" side (toward the electron gun) will also be positive. If there is no positive charge on an area being scanned at a given instant, the electrons in the scanning beam will be repelled at that instant. If there is a positive charge on an area scanned, some electrons will be absorbed, and the rest will be repelled. In this way, the electron beam is modulated, or controlled in intensity, corresponding to light reflected by the portion of the image scanned at that instant.

The repelled electrons, constituting the modulated current, strike a comparatively large electrode around the aperture of the electron gun from which they are shot. Secondary electrons are produced, and these enter several stages of secondary electron multiplication. The output is the so-called **video signal,** and this is ready for amplification in wideband amplifiers, and then for radiation by a local television station, or for transmission over a coaxial cable system, or a radio-relay system, to operate some distant television station.

The Image Orthicon and associated apparatus constitute the **television camera.**[77] The term video is used, technically, to describe the original signal coming from the television pickup device. There is a tendency to use this term to specify television in general. The video band extends from about 30 to 4,500,000 cycles.

Television Transmitters. Television channels are assigned in the very-high-frequency portion of the radio spectrum. For this reason, special tubes and equipment are used.[78] Special antennas must be used because of the high frequencies involved, and the wide band to be transmitted.[79] Amplitude modulation is used for the video, or picture signal, one sideband is suppressed, and the carrier and other sideband are transmitted. Frequency-modulation is used for the audio signal of the speech or music program accompanying the television image transmission. One important function of the transmitting equipment is the generation and transmission of synchronizing impulses needed for "locking together" the television camera and the television receiver.[80]

Television Receivers. Television receivers include equipment for demodulating the received television image signals and for reproducing the image on a **cathode-ray tube** (page 309). Also, equipment is included for demodulating the accompanying speech or music signals and for reproducing the program on a loudspeaker. There are circuits for synchronizing the scanning of the electron beam in the cathode-ray tube in the receiver with the electron beam used in scanning in the

distant television camera.[80] The superheterodyne principle is used in television receivers.

As explained, the electron beam in the cathode-ray tube of the television receiving set scans in synchronism with the beam in the camera tube. The intensity of the beam of the cathode-ray tube is varied in accordance with the instantaneous intensity of the current from the television camera tube. In this way, the television image is reproduced on the end of the cathode-ray tube. In some television sets the image is viewed directly as it is reproduced on the end of the tube, and in others it is viewed indirectly by a suitable optical system.[81] Special receiving antennas are required.[82]

REFERENCES

1. *American Standard Definitions of Electrical Terms.* A.I.E.E., 1941.
2. *Standards on Transmitters and Antennas—Definitions of Terms.* I.R.E. 1942; also, *Standards on Antennas, Modulation Systems, Transmitters.* I.R.E., 1948.
3. Everitt, W. L. *Communication Engineering.* McGraw-Hill Book Co.
4. Kispaugh, A. W., and R. E. Coram. Low Power Radio Transmitters for Broadcasting. *Proc. I.R.E.,* February, 1933, Vol. 21, No. 2.
5. Terman, F. E. *Radio Engineers' Handbook.* McGraw-Hill Book Co.
6. *The A.R.R.L. Antenna Book.* American Radio Relay League.
7. Sterba, E. J., and C. B. Feldman. Transmission Lines for Short-wave Radio Systems. *Proc. I.R.E.,* July, 1932, Vol. 20, No. 7.
8. Ware, L. A., and H. R. Reed. *Communication Circuits,* John Wiley & Sons, Inc.
9. Alford, A., High-frequency Transmission Line Networks. *Elec. Commun.,* January, 1939, Vol. 17, No. 3.
10. Sterba, E. J. Theoretical and Practical Aspects of Directional Transmitting Systems. *Proc. I.R.E.,* July, 1931, Vol. 19, No. 7.
11. Cox, C. R. Phasing Networks for Broadcast Arrays. *Electronics,* June, 1944, Vol. 17, No. 6.
12. *Standards on Radio Receivers—Definitions of Terms.* I.R.E., 1942.
13. Bruce, E. Developments in Short-wave Directive Antennas. *Proc. I.R.E.,* August, 1931.
14. Skilling, H. H. *Fundamentals of Electric Waves.* John Wiley & Sons, Inc.
15. *Standards of Good Engineering Practice Concerning Standard Broadcast Stations.* Federal Communications Commission, 1947.
16. Fink, D. G. Trends in Electronic Engineering. *Elec. Eng.* July, 1948, Vol. 67, No. 7.
17. Mather, R. C., and S. B. Wright. The Compandor—An Aid against Static in Radio Telephony. *Elec. Eng.,* June, 1934, Vol. 53, No. 6.
18. Wright, S. B. The Vodas. *Elec. Eng.,* August, 1937, Vol. 56, No. 8.
19. Blackwell, O. B. Transatlantic Telephony—The Technical Problem. *Bell System Tech. J.,* April, 1928, Vol. 7, No. 2.
20. Jewett, F. B. Transatlantic Telephony. *Sci. Monthly,* August, 1927, Vol. 25.
21. Bown, R. Transoceanic Radiotelephone Developments. *Proc. I.R.E.,* September, 1937, Vol. 25, No. 9.

22. Taylor, C. C. Radio Telephone Noise Reduction by Voice Control at Receiver. *Elec. Eng.*, August, 1937, Vol. 56, No. 8.
23. Polkinghorn, F. A., and N. F. Schlaack. A Single-sideband Short-wave System for Transatlantic Telephony. *Proc. I.R.E.*, July, 1935, Vol. 23, No. 7.
24. Rose, C. F. P. Transoceanic Radio Amplifier. *Bell Lab. Record*, September, 1946, Vol. 24, No. 9.
25. Gabriel, J. C. Single-sideband Short-wave Receiver. *Bell Lab. Record*, November, 1939, Vol. 18, No. 3.
26. Sterba, E. J. Short-wave Transmitting Antenna. *Bell Lab. Record*, August, 1929, Vol. 7, No. 12.
27. Oswald, A. A. Transoceanic Telephone Service—Short-wave Equipment. *Trans. A.I.E.E.*, April, 1930, Vol. 49.
28. Bruce, E., and A. C. Beck. Experiments with Directivity Steering for Fading Reduction. *Proc. I.R.E.*, April, 1935, Vol. 23, No. 4.
29. Friis, H. T., and C. B. Feldman. A Multiple Unit Steerable Antenna for Short-wave Reception. *Proc. I.R.E.*, January, 1934, Vol. 22, No. 1.
30. Beverage, H. H., and H. O. Peterson. Diversity Receiving System of R.C.A. Communications, Inc., for Radiotelegraphy. *Proc. I.R.E.*, April, 1931, Vol. 19, No. 4.
31. Peterson, H. O., H. H. Beverage, and J. B. Moore. Diversity Telephone Receiving System of R.C.A. Communications, Inc. *Proc. I.R.E.*, April, 1931, Vol. 19, No. 4.
32. Heising, R. A. Telephony between Ship and Shore. *Bell Lab. Record*, January, 1930, Vol. 8, No. 5.
33. Wilson, W., and L. Espenschied. Radio Telephone Service to Ships at Sea. *Bell System Tech. J.*, July, 1930, Vol. 9, No. 3.
34. Hansen, E. B. Ship-to-Shore Radio in Puget Sound Area. *Elec. Eng.*, August, 1935, Vol. 54, No. 8.
35. Schlaack, N. F., and F. A. Polkinghorn. An Unattended Ultra-short-wave Radio Telephone System. *Proc. I.R.E.*, November, 1935, Vol. 23, No. 11.
36. McDonald, J. F. Recent Developments in Radiotelegraph Transmitters for Shore Stations. *R.C.A. Rev.*, December, 1947, Vol. 8, No. 4.
37. Sparks, S., and R. G. Kreer. Tape Relay System for Radio-telegraph Operation. *R.C.A. Rev.*, September, 1947, Vol. VIII, No. 3.
38. Sparks, S. International Radiotelegraph Communications. *A.I.E.E. Miscellaneous paper*, No. 48–131. Abstracted in *Elec. Eng.*, June, 1948.
39. Carson, J. R. Notes on Theory of Modulation. *Proc. I.R.E.*, February, 1922, Vol. 10, No. 2.
40. van der Pohl, B. Frequency Modulation. *Proc. I.R.E.*, July, 1930, Vol. 18, No. 7.
41. Roder, H. Amplitude, Phase, and Frequency Modulation. *Proc. I.R.E.*, December, 1931, Vol. 19, No. 12.
42. Armstrong, E. H. A Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation. *Proc. I.R.E.*, May, 1936, Vol. 24, No. 5.
43. Everitt, W. L. Frequency Modulation. *Elec. Eng.*, November, 1940, Vol. 59, No. 11.
44. Travis, C. Automatic Frequency Control. *Proc. I.R.E.*, October, 1935, Volume 23, No. 10.

45. Foster, D. E., and S. W. Seeley. Automatic Tuning, Simplified Circuits, and Design Practice. *Proc. I.R.E.,* March, 1937, Vol. 25, No. 3.
46. Crosby, M. G. Reactance-tube Frequency Modulators. *QST,* June, 1940; also *R.C.A. Rev.,* June, 1940, Vol. 5, No. 1.
47. Morrison, J. F. A New Broadcast-transmitter Circuit Design for Frequency Modulation. *Proc. I.R.E.,* October, 1940, Vol. 28, No. 10.
48. Adler, R. A New System of Frequency Modulation. *Proc. I.R.E.,* January, 1947, Vol. 35, No. 1.
49. Wrathall, L. R. Frequency Modulation by Non-linear Coils. *Bell Lab. Record,* March, 1946, Vol. 25, No. 3.
50. Seeley, S. W., and J. Avins. The Ratio Detector. *R.C.A. Rev.,* June, 1947, Vol. VIII, No. 2.
51. *Standards of Good Engineering Practice Concerning FM Broadcast Stations.* Federal Communications Commission, 1945.
52. Fremodyne F-M Receivers. *Electronics,* January, 1948, Vol. 21, No. 1.
53. Roberts, W. van B. Input Impedance of a Folded Dipole. *R.C.A. Rev.,* June, 1947, Vol. 8, No. 2.
54. Bailey, A. Mobile Radio Communication. *Trans. A.I.E.E.,* Vol. 67, 1948.
55. Peterson, Jr., A. C. Vehicle Radiotelegraphy Becomes a Bell System Practice. *Bell Lab. Record,* April, 1947, Vol. 25, No. 4.
56. Chu, L. J., and W. L. Barrow. Electromagnetic Horn Design. *Trans. A.I.E.E.,* Vol. 58, 1939.
57. Cutler, C. C. Parabolic-antenna Design for Microwaves. *Proc. I.R.E.,* November, 1947, Vol. 35, No. 11.
58. Kock, W. E. Metallic Delay Lenses. *Bell System Tech. J.,* January, 1948, Vol. 27, No. 1.
59. Ullrich, E. H. Ultra-short Wave Communication. *Elec. Commun.,* July, 1937, Vol. 16, No. 1.
60. Labin, E. Microwave Radio Relay Systems. *Elec. Commun.,* June, 1947, Vol. 24, No. 2.
61. Thompson, L. E. A Microwave Relay System. *Proc. I.R.E.,* December, 1946, Vol. 34, No. 12.
62. Friis, H. T. Microwave Repeater Research. *Bell System Tech. J.,* April, 1948, Vol. 27, No. 2.
63. Durkee, A. L. A Microwave Relay System Between New York and Boston. *Bell Lab. Record,* December, 1947, Vol. 25, No. 12.
64. Chaffee, J. G. Terminals for the New York–Boston Radio Relay System. *Bell Lab. Record,* March, 1948, Vol. 26, No. 3.
65. Brotherton, M. Making Energy Talk. *Bell Lab. Record,* June, 1948, Vol. 26, No. 6.
66. Grieg, D. D. Pulse-Count Modulation. *Elec. Commun.,* September, 1947, Vol. 24, No. 3.
67. Clavier, A. G., P. F. Panter, and D. D. Grieg. Distortion in Pulse-Count Modulation System. *Elec. Eng.,* November, 1947, Vol. 66, No. 7.
68. Goodall, W. M. Telephony by Pulse-Code Modulation, *Bell System Tech. J.,* July, 1947, Vol. 26, No. 3.
68(*a*). Feldman, C. B. A 96-channel Pulse-Code Modulation System. *Bell Lab. Record,* September, 1948, Vol. 26, No. 9.
69. Grieg, D. D., and A. M. Levine. Pulse-time-modulation Multiplex Radio

Relay System-Terminal Equipment. *Elec. Commun.*, June, 1946, Vol. 23, No. 2.
70. Meacham, L. A., E. Peterson, and R. W. Sears. Pulse-Code Modulation Using Electron Beam Tubes. *Bell System Tech. J.*, January, 1948, Vol. 27, No. 1.
71. Oliver, B. M., J. R. Pierce, and C. E. Shannon. Philosophy of PCM. *Proc. I.R.E.*, November, 1948, Vol. 36, No. 11.
72. Ives, H. E., J. W. Horton, R. D. Parker, and A. B. Clark. The Transmission of Pictures over Telephone Lines. *Bell System Tech. J.*, April, 1925, Vol. 4, No. 2.
73. Reynolds, F. W. New Telephotograph System. *Bell System Tech. J.*, October, 1936, Vol. 15, No. 4.
74. Two-way Television: A group of three papers on television. *Bell System Tech. J.*, July, 1930, Vol. 9, No. 3.
75. Ives, H. E., and A. L. Johnsrud. Television in Colors by a Beam Scanning Method. *J. Optical Soc. Am.*, January, 1930, Vol. 20.
76. Rose, A., P. K. Weimer, and H. B. Law. The Image Orthicon—A Sensitive Television Pickup Tube. *Proc. I.R.E.*, July, 1946, Vol. 34, No. 7.
77. Kell, R. D., and G. C. Sziklia. Image Orthicon Camera. *R.C.A. Rev.*, March, 1946, Vol. 7, No. 1.
78. Downie, J. W., L. M. Ewing, H. B. Fancher, and J. E. Keister. Design of Television Transmitters for Low Level Modulation. *Tele-Tech,* December, 1947, Vol. 6, No. 12.
79. FM and Television Antennas. *Electronics,* November, 1948, Vol. 21, No. 11.
80. Fink, D. G. *Principles of Television Engineering.* McGraw-Hill Book Co.
81. Wright, A. Television Receivers, *R.C.A. Rev.*, March, 1947, Vol. 8, No. 1.
82. Kolster, F. A. Antenna Design for Television and FM Reception. *Proc. I.R.E.*, October, 1948, Vol. 36, No. 10.

REVIEW QUESTIONS

1. Explain what is meant by frequency translation, and discuss several applications in radio.
2. What general classification is made for the systems of modulation used in radio?
3. What is the purpose of the buffer amplifier shown in Fig. 1?
4. What is the difference between high-level and low-level modulation?
5. What is the source of the sideband power in the modulating circuits of Figs. 3 and 4?
6. If the carrier component is suppressed in a radio system, how is reception possible?
7. Name several important reasons why tetrodes and pentodes sometimes are used as modulators.
8. What are meant by the terms resonant and non-resonant feeders?
9. One of the circuits of Fig. 5 is for connecting at a low-impedance point, and the other is for connecting at a high-impedance point. Explain what this means.
10. If a half-wave antenna is fed at the end, will the input impedance be greater or less than if it is fed at the center? Why?

11. Explain why a thermomilliammeter indicates when it is moved along one wire as discussed on page 496.
12. Fully explain why the matching section of page 497 is exactly $\frac{1}{4}\lambda$ in length.
13. Why are phase-shifting networks sometimes used with broadcast antenna feeders?
14. Enumerate several advantages of the superheterodyne radio receiver over other types.
15. What is the approximate bandwidth required by an amplitude-modulation broadcast transmitter? By a frequency-modulation broadcast transmitter?
16. Should the receiving antenna of a broadcast receiver be directional or nondirectional?
17. What is the purpose of tuning preceding the converter in a superheterodyne receiver?
18. Explain the difference between primary and secondary service areas.
19. Explain how it is possible to send and receive on the same radio frequency.
20. Discuss the advantages of the rhombic antenna over comparable types.
21. How do diversity systems operate?
22. Explain the points of similarity and difference between phase and frequency modulation.
23. Why is a beating oscillator necessary in a radio-telegraph receiver?
24. How would a radio-broadcast receiver be designed if amplitude-modulation broadcasting were by a single sideband?
25. Under what conditions will a phase modulator produce a frequency-modulated wave?
26. For comparable conditions, will a phase-modulated signal or a frequency-modulated signal require the greater bandwidth? Why?
27. Why is a balanced modulator used in Fig. 24?
28. What is the purpose of a pre-emphasis circuit?
29. Does the Phasitron alone produce phase or frequency modulation?
30. What is the purpose of resistor R_2 of Fig. 27? What is a typical value?
31. In terms of percentage, what frequency drift is permitted in frequency-modulation broadcast transmitters?
32. Briefly compare the coaxial cable with radio-relay systems.
33. In the New York–Boston radio-relay system why is the signal reduced to 65 megacycles at each repeater? Why is it received and transmitted at slightly different frequencies?
34. What is meant by pulse-code modulation? What is an important advantage?
35. Briefly discuss the operation of the Image Orthicon tube.

PROBLEMS

1. A modulated amplifier that is operated in class C with the modulating signal injected into the plate circuit has an unmodulated carrier output of 5.0 kilowatts. With 100 per cent sinusoidal modulation, what is the approximate direct power drawn by the plate circuit, the approximate alternating power drawn from the modulating amplifier, the approximate direct power drawn by the modulating amplifier, and the approximate total 60-cycle alternating power drawn by the transmitter.
2. If the power-output tubes of a transmitter should work at 1825 kilocycles into 3250 ohms resistance and if the input impedance to the transmission line is 175 ohms resistance, design a simple impedance-matching network for coup-

ling the transmitter to the line. Repeat if the line impedance is 1250 ohms resistance.

3. In the system of Fig. 7, the standing-wave ratio is 4.7 without the stub. If the frequency is 59.75 megacycles, determine the lengths and locations of open and shorted stubs to remove the standing waves. Is it possible to substitute for the open stub a shorted stub that is $\frac{1}{4}\lambda$ longer than the open stub? Why?
4. The length of the antenna in Fig. 9 is 0.5 λ. The feeders are of No. 10 wire, 8 inches apart. Two copper tubes of outside diameter 1.1 inch are available. Can they be used to approximately match the impedances? If so, what spacing must they have? If not, what is a suggested diameter?
5. A vertical broadcast antenna with a base insulator is fed by an underground coaxial cable. What lightning protection would you recommend? If the tower is struck, can the transmitter cause a radio-frequency power arc on the system? If so, where will it probably occur?
6. The functioning of the 1.0-megohm resistor and 150-micromicrofarad capacitor of Fig. 14 is explained on page 501. Assume that the station being received is modulated with a 1000-cycle tone and that the circuit is in a typical superheterodyne and calculate the impedance offered to the principal alternating components that will be impressed on this parallel combination.
7. In discussing Fig. 26 it is explained that sometimes the capacitor between the primary and the secondary is omitted and a coil is used. Assume that this change and other necessary modifications have been made and draw the required vectors. Prepare a discussion to explain the operation of the circuit.
8. On page 525 the statement is made that the frequency multipliers increase both the frequency and the frequency deviation. Prove this to be true.

CHAPTER 14

INTERFERENCE AND NOISE

Introduction. In the past, noise in communication systems has been considered objectionable largely because of its interfering effects, and for similar reasons. It is now known, however, that there is a relationship between the amount of noise in a circuit and the bandwidth required to transmit information.

Bandwidth and Noise in Communication Systems. Of much importance in communication is the bandwidth required to pass information. In 1928 Hartley stated[1] that "the total amount of information which may be transmitted over such a system is proportional to the product of the frequency-range which it transmits by the time during which it is available for transmission." The system referred to was one having a restricted transmission range, a characteristic common to most communication systems used in practice. This principle became known as the **Hartley law,** and was accepted for 20 years. The law in itself apparently is valid.

With the development of pulse modulation, bandwidth requirements were reinvestigated by several independent groups[2, 3] who found that the amount of interfering circuit noise in a system was also a factor determining bandwidth requirements. This has resulted in a modified law, which has been stated in several ways. Expressed mathematically,[2]

$$C = w \log_2(1 + S/N), \tag{1}$$

where C is the capacity of the channel to carry information per unit time, w is the bandwidth of the channel, and S/N is the signal-to-noise ratio in *power units.*

This law (which is sometimes referred to as the modified Hartley law) has very important implications. Thus, if the signal-to-noise *power* ratio is always maintained high, the bandwidth requirements may be low. Hence, in telephone, telegraph, or television transmission systems, if the power is always maintained high, less bandwidth is required by each signal for the same quality of transmission. Either more signals, or higher quality, are possible within a given bandwidth if the signal-to-noise power ratio is kept high.[4]

It has been pointed out[2] that the noise-reducing property of frequency modulation is in accordance with this principle. Thus, if an amplitude-modulated radio transmitter were modulated with an audio signal having a maximum frequency of 15,000 cycles and if *both* sidebands were transmitted, the bandwidth required would be 30,000 cycles. If a frequency-modulation radio transmitter of the broadcast type were modulated with the same signal, the bandwidth required would be about 200,000 cycles wide. Hence, for comparable power conditions, a frequency-modulation station should give a considerable improvement in received signal-to-noise ratio. It appears that pulse-time modulation and particularly pulse-code modulation will be able to utilize the advantages indicated by this principle.[4]

Interference and Noise in Wire Communication Systems. Interference will be considered in a broad sense and will include those *electrical* effects which disturb, or handicap, in any way normal use and operation. Interference may result in the operation of protective equipment, such as fuses, or may result in damage to the physical plant if the protection is inadequate or improper. One of the most serious types of interference is **noise,** defined[5] as "any extraneous sound tending to interfere with the proper and easy perception of those sounds which it is desired to receive." Because acoustic noise was considered in Chapter 2, the present discussion will be limited to **circuit noise,** defined[5] as "noise which is brought to the receiver electrically from a telephone system, excluding noise picked up acoustically by the telephone transmitters."

Sources of interference may be classified as natural or artificial. The natural classification includes lightning, static, the aurora borealis phenomena, and dust storms phenomena. Artificial sources include power lines, electric railway systems, and other communication circuits. Induced currents from these and other sources cause noise in the telephone receivers. An unusual source of noise is explained in reference 6.

Telephone lines are designed for, and operated with, low voltages and currents and must therefore be protected not only from the induced noise currents just considered but also from high-voltage natural and artificial hazards as well. Lightning offers the greatest natural hazard and, through either direct or induced strokes, can injure or destroy telephone equipment. Heavy earth currents that accompany magnetic storms and auroral displays have damaged grounded apparatus and cables. Power lines, including electric trolleys and feeders, offer the greatest artificial hazards. These lines may come in direct contact with telephone circuits, or may *induce* relatively high

voltages into these circuits as this chapter will explain. Electrolysis also is a source of considerable trouble.

Effects of Interference. Interference from paralleling power and electric traction systems commonly is **circuit noise,** that is, a buzzing or humming in the telephone receiver. If the interference is between two paralleling telephone lines, it is called **crosstalk.** If one telegraph circuit induces signals into a paralleling telegraph circuit, it is termed **crossfire.**

Crosstalk is usually easily controlled because, *first,* it occurs between circuits operating at about the same power level, and *second,* it is usually between lines of the same system, engineered and operated by the same organization. If crosstalk is noticeable, it not only interferes with telephone service but also destroys secrecy. Crosstalk interferes with normal telephone conversation by diverting attention from the conversation in progress to the one heard through induction from another circuit. There is, furthermore, a masking or interfering effect of considerable importance.

Induction from paralleling power lines offers a serious problem. One reason for this is the great difference in the voltage, current, and power magnitudes involved. If but an extremely small amount of the power in a power system is induced into a low-level communication circuit, it may be of the same order of magnitude as the useful telephone currents.

Induction from power systems may be of two types: first, **low-frequency induction;** and second, **noise-frequency induction.** (References 7 to 13 inclusive.) Low-frequency induction includes only induction at the *fundamental frequency* of the power system (usually 60 cycles), and ordinarily is bothersome only at times of abnormal power-line conditions. Noise-frequency induction is, however, due to the *harmonics* of the fundamental power frequency and is usually present at all times, although its effect is more pronounced under abnormal power-line conditions.

Both low-frequency and noise-frequency induction interfere with telephone service. Among the effects produced by low-frequency induction are[11] service interruptions, false signals, telegraph signal distortion, damage to plant, electric shock, and acoustic shock. Regarding this last effect, it is interesting to note that copper oxide varistors sometimes are connected across operators' telephone sets to reduce acoustic shock; also, that the non-resonant telephone receiver produces much less noise when a transient current passes through it. Noise currents tend to mask the useful speech or signal currents. A noisy circuit having a certain loss is equivalent from a telephone

standpoint to a more quiet circuit having a *greater loss*. The added losses due to noise are known as **noise transmission impairments,** abbreviated **N.T.I.** Much more speech power must be supplied to the telephone user to override the induced and room noises.

History of Inductive Coordination. Early in the history of the electrical industry, the telegraph companies operated alone in the field. Interference between adjacent telegraph circuits was experienced but was not serious. Earth currents, set up during displays of the aurora borealis,[14] affected the grounded lines seriously at times.

The early telephone sets were operated over grounded lines, and interference between telephone and telegraph lines, and between the telephone circuits themselves, was immediately experienced. Power systems and electric railways were introduced during this same period; and, especially after alternating current came into use, much interference resulted.

The history of interference is interesting[15, 16] as the following quotation indicates.

> About 1883 a new iron wire was made which, it was claimed, would be free from inductive disturbances. Its cross-section was shaped like a four-leaf clover and the grooves ran around the wire in a spiral. The company that was making and promoting this wire interested themselves in its use on more than a dozen experimental lines. They had a theory that the voice current would follow the spiral and in some way produce a beneficial effect. An officer of the spiral wire company said it was not understood by electricians.*

The use of two-wire, or metallic, instead of grounded telephone circuits did much to reduce noise. They were introduced commercially by Carty.[16] Disturbances were further reduced by transpositions explained by Carty in his paper[17] presented in 1891.

The process by which electric energy is transferred from one circuit to another, causing noise or other disturbances in low-level communication circuits, is termed **induction.** The interference caused thereby is called **inductive interference.** The transfer of energy may be largely through the effect of the electric field or the magnetic field, or by both. Power engineers and communication engineers cooperate in minimizing these effects, such activities leading to **inductive coordination.**

Perhaps the earliest comprehensive cooperative survey was begun in 1912. Rules were formulated and published[18] for the coordination of power and telephone systems so that inductive disturbances would be minimum. After 1921, coordination activities were centered in a

* Reprinted by permission from *Beginnings of Telephony* by F. L. Rhodes, Harper & Bros.

joint general committee of the National Electric Light Association and the Bell Telephone System.[19] More recently, committees of the Edison Electric Institute and the Bell System have continued this work. Excellent cooperation exists between the power and communication organizations, and inductive interference has been rendered controllable.

Factors Determining Interference.[20, 21] The three important factors[11] that determine the inductive interference experienced under given conditions are as follows: (1) influence factors; (2) susceptiveness factors; (3) coupling factors.

Each power circuit has certain distinguishing characteristics of its line and connected apparatus, such as generators, transformers, and loads. These determine the **influence factor.** Similarly, each tele-

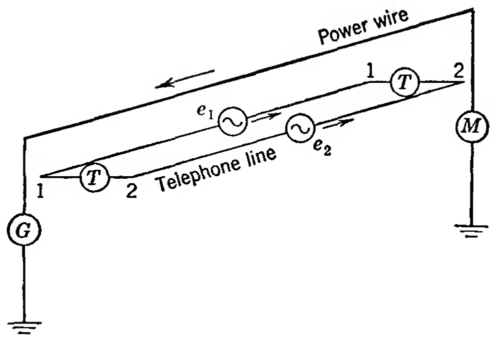


FIG. 1. A parallel consisting of a grounded power line (fed by generator G and connected to motor M) and a two-wire, or metallic, telephone line. The reason for considering a grounded power line will be explained on page 557.

phone circuit has special features, and these determine its responsiveness to the electric and magnetic fields from the power line, and the **susceptiveness factor.** Also, each parallel has individual characteristics, such as distance between the power and telephone lines and length of parallel. These determine the **coupling factor.** It is emphasized that these factors and the discussions that follow are general in nature; they apply to electrical noise pickup by a radio or sound system.

Methods of Induction. Electric energy is transferred from power lines into communication circuits and from one communication circuit into another in three ways. *First,* energy transfer may occur through leakage paths between the two circuits. This is usually of little importance, and its solution is obvious.

The magnetic field offers the *second* medium for introducing interfering currents into communication circuits. Thus, in Fig. 1, assume that a grounded alternating-current power line parallels a metallic two-wire telephone line as indicated. The power-line current will then produce a magnetic field about the power wire, and this field will link with the telephone line as shown in Fig. 2. Instantaneous induced voltages, represented by the generators e_1 and e_2, will be produced in the telephone wires. Since wire 1 is closer to the power line than wire 2, wire 1 will have a greater voltage induced in it by the alternating magnetic field, and thus a *difference of potential* e_1-e_2 will exist between wires 1 and 2. This difference of potential (often called **metallic-circuit induction,** and sometimes **transverse induction**) will cause a current to flow in the telephone circuit, and noise from the power line will accordingly be heard in the telephone receivers.

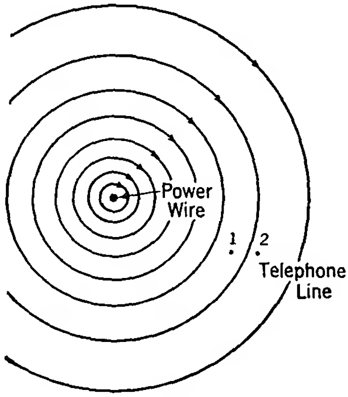


Fig. 2. The magnetic field of the power line of Fig. 1 links with the telephone conductors.

The magnetic field causes a voltage to be induced *in series* with each wire. As Fig. 3 indicates, one end of each wire is positive and the other is negative. A difference of potential (often called **longitudinal-circuit induction**) will also exist between the wires and ground. If a high-impedance voltmeter V is connected as in Fig. 3, the voltmeter will deflect because the circuit is completed through the distributed capacitance represented by C_{g1}.

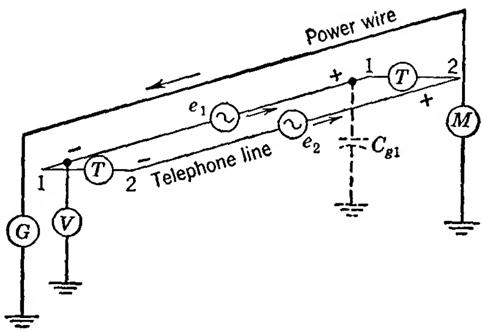


Fig. 3. The magnetic field of Fig. 2 induces voltages in each telephone wire as indicated by e_1 and e_2. Since e_1 is greater than e_2, noise currents will be forced around through the telephone sets. These induced voltages tend to act between the telephone wires and ground as indicated by the fact that voltmeter V will indicate the presence of a voltage. This voltage to ground will force currents through capacitive and leakage paths to ground.

The *third* method by which noise energy is introduced into paralleling telephone circuits can be

explained from a consideration of the power-line voltages and distributed capacitances between the circuits involved (Fig. 4). Considering for the moment one wire only, suppose that C_1 represents the distributed capacitance between the power wire and wire 1, and that C_{g1} represents the capacitance of wire 1 to ground. The effective voltage E between the power wire and ground will divide inversely proportional to the capacitances, and thus $E_{g1} = EC_1/(C_1 + C_{g1})$. The voltage from wire 2 to ground will be less than E_{g1} since the respective distances between wires are greater, but the capacitances to ground are the same.

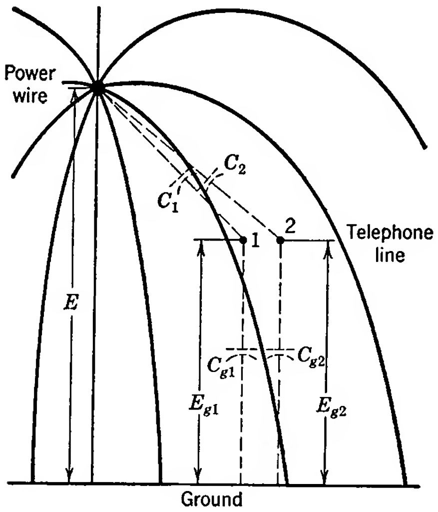


FIG. 4. The power-line voltage E to ground acting through the distributed capacitance raises the telephone line wires to *unequal* voltages above ground. The voltage E_{g1} exceeds E_{g2}. The shape of the electric field is approximately as indicated.

Since the two voltages to ground are unequal, the wires will have a difference of potential between them, and this will cause noise currents to flow through the connected telephone sets. The way in which the two telephone wires are raised to a voltage above ground is also shown in Fig. 5. The induced voltages are represented by the small generators E_{g1} and E_{g2}.

From this section it is apparent that the power-line *current* and the resulting magnetic field induce voltages which act as small generators connected *in series* in the line wires (Fig. 1). Also, the power-line *voltage* and the resulting electric field raise the paralleling telephone-line wires to voltages above ground, and these voltages act as small generators connected *between the telephone wires and ground* (Fig. 5).

In this discussion and the one which follows, it has been assumed that the telephone line is paralleled by a power line consisting of a single grounded wire. Such lines seldom exist for commercial power purposes, but, as will be discussed later, under certain conditions a three-phase power line has the characteristics of a grounded circuit as

just considered. Also, the grounded system applies for the single trolley wire of an electric railway.

Line Unbalances. In the preceding section it was shown that induced voltages were of two types: *first,* voltages acting in series in each wire of the disturbed circuits; and *second,* voltages acting between the wires and ground.

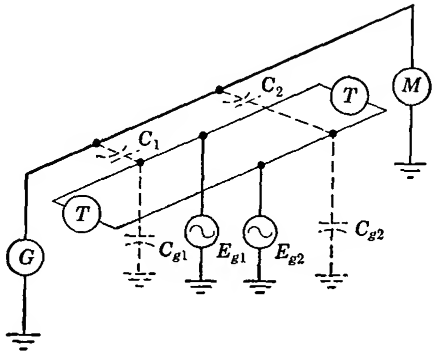


FIG. 5. Showing Fig. 4 somewhat in perspective. The power-line voltage acting through the distributed capacitances raises the telephone wires to *unequal* voltages above ground as shown by the small generators. The voltage E_{g1} exceeds E_{g2}. The internal impedance of each generator consists largely of the capacitance between the wire and ground.

As was also shown in the preceding discussion, if *either* the voltages induced in series or those to ground *are not equal,* then noise currents will flow through the connected telephone sets. In the following paragraphs it will be assumed that the telephone wires are raised to *equal voltages above ground,* and the way these *equal* voltages cause noise will be considered. Later in the section the effect of *equal series voltages* will be discussed.

Suppose that the two telephone wires of Fig. 6 are raised to *equal* potentials E_{g1} and E_{g2} above ground, that the series impedances Z_1 and Z_2 of each section of line wire are exactly *the same,*

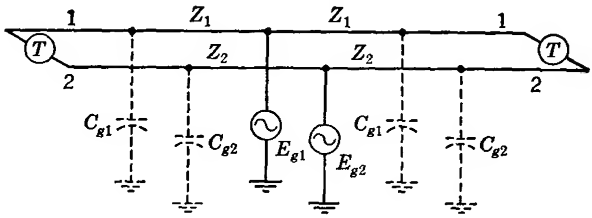


FIG. 6. An approximate simplified circuit for a balanced telephone line on which *equal* voltages E_{g1} and E_{g2} exist to ground. The series impedance Z of each elemental section is the same. Also, the impedance to ground of each elemental section is the same. In these simplified circuits the shunt elements (Fig. 3, page 196) and leakage paths to ground are not shown.

and that the impedances to ground also are *identical.* Then the telephone line is **balanced,** and the *equal* voltages *to ground* will produce

no noise currents in the telephone receivers. In commercial systems it is impractical, if not impossible, to keep the telephone lines exactly balanced, and thus *equal* voltages induced to ground *will* produce noise currents in the connected telephone sets.

Figure 7 shows a line with a **series unbalance,** such as would be caused by a poor joint in the telephone line wire. The capacitance to ground of each elemental section is represented approximately by

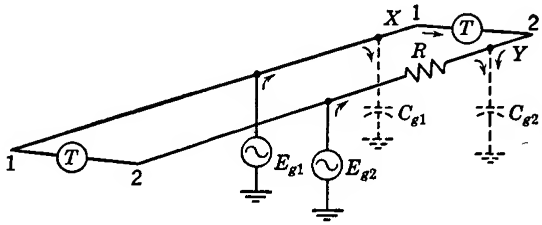


FIG. 7. Illustrating the way in which a series unbalance, such as a poor splice, causes noise currents to flow when the telephone wires are raised to *equal* voltages E_{g1} and E_{g2} above ground by a power parallel. The telephone wires are assumed to be balanced to ground. Points X and Y will not be at the same potential, and a noise current will flow through the connected telephone sets.

the lumped capacitances C_{g1} and C_{g2}. The *equal* voltages to ground E_{g1} and E_{g2} acting on the telephone wires send currents to ground through these capacitances. But the current flowing in wire 2 must flow through the unbalance R, and the two points X and Y will accordingly be at different potentials. As a result, noise currents will flow through the telephone sets.

Unbalances to ground originate in many ways. The most common causes are tree "leaks," defective or dirty insulators, and poorly designed or maintained terminal equipment. An **unbalance to ground** causes noise as shown in Fig. 8; the "leak" R_g is assumed to be of high resistance. With this arrangement point Y is at a lower potential than point X, and noise currents will flow through the connected telephone sets even when the voltages induced to ground, E_{g1} and E_{g2}, are equal.

In addition to causing noise, a high induced voltage to ground may offer a serious electrical hazard to employees and to equipment. When a telephone line is very close to a power line, the telephone line may assume potentials of hundreds or even thousands of volts above ground. Voltages to ground on a telephone line necessitate a high quality of maintenance of both lines and equipment if excessive noise is to be prevented. Thus, even if the noise is not actually produced,

the *presence* of the induced voltage necessitates higher maintenance expenditures. In many instances excessive noise on telephone circuits has been reduced when the balance of the lines, both series and to ground, was improved by removing faulty line joints and by replacing broken insulators and removing tree and brush "leaks."

Leaving now the effect of voltages to ground, the effect of equal *series* voltages will be briefly considered. In Fig. 3 it was shown that voltages induced in series would act to ground. Thus, even if the induced voltages are equal, unbalances either in series or to ground

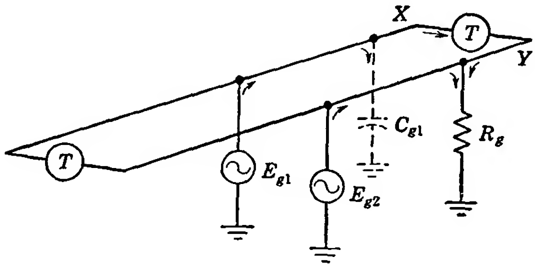


FIG. 8. Illustrating the way in which an unbalance to ground, such as would result from a poor insulator or a tree "leak," causes noise currents to flow. The two voltages E_{g1} and E_{g2} to which the telephone line is raised above ground are *equal.* The resistance of the ground is assumed negligible. The potential of point Y approaches as a limit that of ground, and noise currents accordingly flow through the telephone set. The series impedances of the two telephone wires are assumed equal.

may cause the currents which flow to be unequal. Then, the ends of the two line wires to which the telephone sets are connected may be at different potentials, and noise currents may flow.

Telephone Line Transpositions. It has been shown that two types of voltages—voltages *in series* in the line wires, and voltages *between the wires and ground*—may be induced in telephone circuits by paralleling electric power lines. The effects on these voltages of transpositions in the *telephone lines* will now be considered. Transpositions in the power circuits will be considered in later sections.

The circuit of Fig. 1 has been redrawn in Fig. 9 with a transposition at the center of the telephone line; this balances the telephone line with respect to the power wire. Voltages will still be induced in each telephone wire by the varying magnetic field; these are indicated by e_1, e_1', and e_2 and e_2' of Fig. 9. Since the two telephone wires now occupy the same relative positions, $e_1 + e_1' = e_2 + e_2'$, and, since these sums are substantially equal and the voltages are opposite, they

will almost neutralize each other. Telephone line transpositions therefore *tend to equalize the voltages induced in series in each line wire by magnetic coupling.* This often greatly reduces the noise.

The effect of transpositions on induction from power-line voltages through the electric field will now be considered. As shown in Figs. 4, 5, and the accompanying discussions, this field tends to raise the telephone wires to unequal voltages above ground.

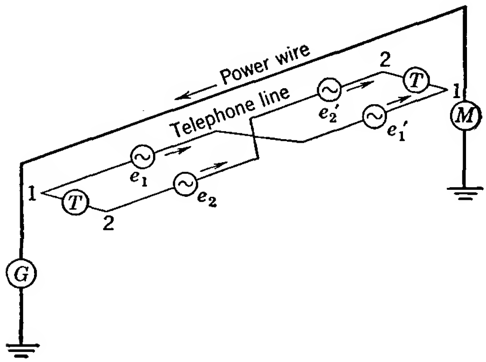


FIG. 9. A transposition in the telephone line balances it with respect to the power line. With this arrangement, $e_1 + e_1' = e_2 + e_2'$, and negligible resultant voltage will exist to force a noise-producing current through the connected telephone sets.

With the circuit transposed, conditions change from those of Fig. 5 to those of Fig. 10. In this figure, voltage $E_{g1} = E_{g2}'$ and $E_{g2} = E_{g1}'$, but the two former exceed the two latter in magnitude. These voltages at a given instant are assumed to act in the direction shown. The equivalent circuit is also shown in Fig. 10, and from this the effect of transpositions can be explained.

Analyzing the equivalent circuit, it will be found that the voltage $E_{g2}'-E_{g1}'$ will tend to send a noise current I' up through the telephone at the right, and that the voltage $E_{g1}-E_{g2}$ tends to send a current I *down* through the same telephone set. This current I will be less than I' because of the series impedances in the line wires. Hence, a resultant noise current will flow up through the telephone set at the right. Thus, a difference of potential, determined by the magnitude of the line series impedances (other facts being comparable), exists between the ends of the two wires at the right (and at the left as well).

If the transposition sections are shortened, the series line impedance per section will be reduced, and the difference of potential across the telephone sets will be lessened. Although the transpositions reduce the noise-producing voltage between the wires as just explained, they do *not* appreciably affect the voltages to ground which may, therefore, act through unbalances to cause noise.

In both the explanations just given, it was pointed out that transpositions tend to reduce noise. It has sometimes been thought that

one transposition located in the center of a disturbed circuit of *any* length will reduce the noise to a very low level. (In fact, such an arrangement was tried without success on the early New York–Philadelphia circuits.)

One transposition is not effective for several other reasons. The *first* is that because of line attenuation the magnitudes of the current and voltage are not exactly the same at each point along the power wires. This will prevent complete equalization of the voltages. Although this effect may be negligible for many power systems, it exists nevertheless. It is especially important when considering crosstalk from telephone circuits which have comparatively high attenuation.

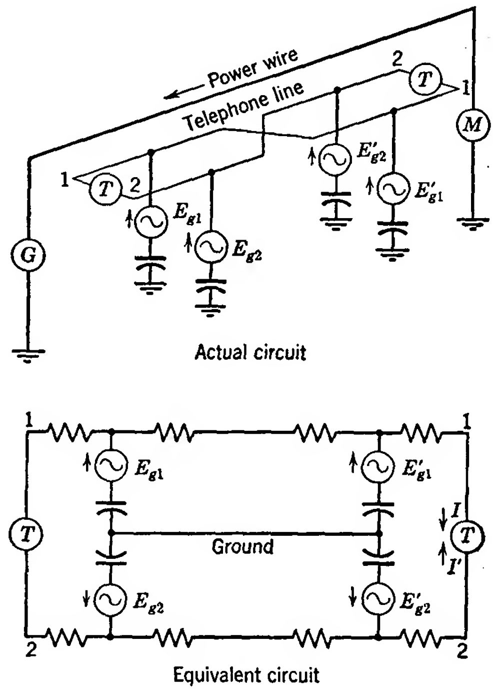


FIG. 10. Actual and equivalent circuits showing how telephone line transpositions reduce noise caused by the telephone line wires being raised to unequal voltages above ground.

The *second* important reason that very long transposition sections are not effective is due to phase relations. As was shown in Chapter 6, a finite time is required for the propagation of current and voltage impulses, and, although these might have the same magnitude (neglecting attenuation) in different parts of a circuit, they would have different phase relations. It is apparent that, if the phase relations are greatly different in two sections of a line, it would be possible to have the induced voltages tending to add. Phase relations are particularly important for the higher frequencies (shorter wavelengths) encountered in communication circuits; it also becomes of increasing importance for the higher power-line harmonics which are in reality the components causing noise.

The practical solution is to have the telephone transpositions (and also the power transpositions as will be shown later) close together, so that the distance between transpositions is but a small fraction of

a wavelength of the impulses in the disturbing circuit. Also, keeping the wires of the disturbing circuit close together will tend to reduce the stray fields and the *influence factor*, and, similarly, keeping the wires of the disturbed circuit close together will tend to reduce the *susceptiveness factor*.

Telephone Transposition Schemes.[22, 23, 24] A very elementary transposition scheme for crosstalk alone is shown in Fig. 11. A study of this diagram will indicate that, if the distance *d* is small, about one mile or less, each circuit will be inductively balanced with respect to each of the others, and thus excessive crosstalk will not occur.

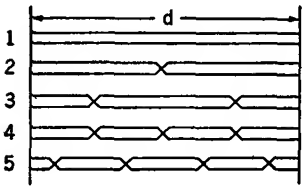


FIG. 11. An elementary transposition scheme for reducing crosstalk.

It is apparent that where many wires are involved, as in a four-arm line having 40 wires, transposition schemes become very complex. This is further complicated by the use of phantom circuits (page 224), for not only must crosstalk between side circuits be minimized but also crosstalk caused by phantoms must be reduced.

From these discussions it should not be inferred that special transposition schemes are designed for each new line. In fact, quite the opposite is true; standard transposition systems[23] are available and are used in new line construction. A standard scheme is shown in Fig. 12.

Attention is called to the four different types of phantom transpositions indicated in the diagram. These, it will be observed, depend on the side-circuit transpositions at the phantom transposition points. When a new line is laid out, transposition schemes such as Fig. 12 are fitted in so that they best meet conditions, especially from the standpoint of coordination with existing or proposed power parallels. The complexity of phantom transpositions is evident from Fig. 12. In fact, the necessity for transpositions is *one* factor preventing the use of "phantoms on phantoms," often called "superphantoms" or "ghosts."

Classification of Power Parallels. From the standpoint of inductive interference there are many possible types of power parallels. The arrangement of the conductors or **configuration** greatly affects the induction.[18] In the following pages, the discussion will be limited to three-phase circuits having triangular configuration. These circuits may be classified as follows: (1) non-transposed, non-grounded; (2) non-transposed, grounded; (3) transposed, non-grounded; and (4) transposed, grounded.

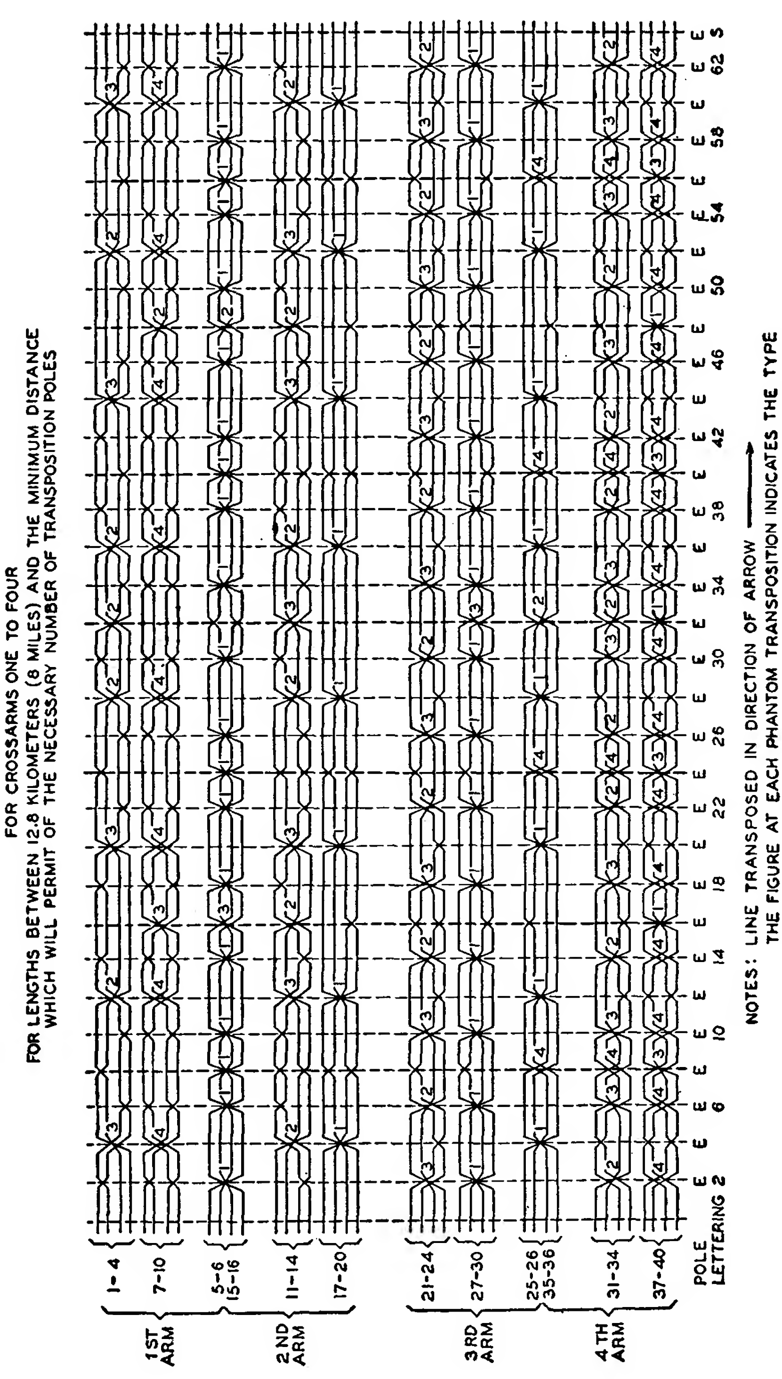


FIG. 12. A standard transposition section, known as the Type E section. Both the side-circuit and phantom transpositions are shown. (Courtesy Bell Telephone System.)

Power lines, when grounded, are usually grounded at the neutrals of three-phase, wye-connected transformer banks. The neutrals of alternators are often grounded. There are in general three types of grounds in common use in this country: *first,* a direct connection to earth; *second,* grounds through current-limiting resistors or reactors; and *third,* grounds through parallel-resonant circuits offering low impedance to the fundamental, but greatly attenuating noise-producing harmonics. The influence factor varies greatly with grounding arrangements.

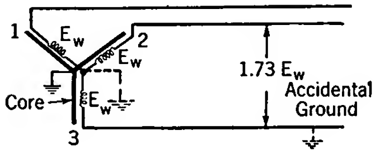


FIG. 13. The effect of an accidental ground on a three-phase system.

Although there are other reasons why it is desirable to ground neutrals, one is that it offers a protection to equipment in the event of a ground on one of the line wires, and thus insulation costs are less. This is shown in Fig. 13. The transformer or alternator windings must be insulated from the core which is assumed to be at ground potential. If the windings are not grounded and an accidental ground occurs on wire 3, the voltages between windings and core at points 1 and 2 will now be approximately $1.73E_w$, where E_w is the voltage in each winding. If, however, a ground is in existence at the neutral of the winding as shown dotted and if then an accidental ground occurs on wire 3, the voltage at points 1 and 2 cannot rise above E_w.

Balanced and Residual Currents and Voltages. In studying inductive interference problems, it is helpful to divide the power-system currents and voltages into components with respect to earth as a reference point. In defining these, it will be well to quote from the California Joint Committee's report.[18] According to this committee, there are

two general classes: (1) "balanced" with respect to earth as a neutral conductor or point of reference, and (2) "residual," completely unbalanced with respect to the earth, i.e., employing the metallic power-circuit conductors, as a group, for one "side" and the earth as the other side of their circuit.*

To quote this committee further,

"Balanced" current components in the several conductors of a power circuit are such that at every instant their algebraic sum is zero. The algebraic sum of the total currents in the several conductors of a power circuit at any instant is the "residual" current. Similarly, the "balanced" voltages of the several conductors are such that their algebraic sum is zero at every instant, while the algebraic sum of the total voltages to ground at any instant is the "residual" voltage. [AUTHOR'S NOTE: The word *algebraic* as used probably implies what is usually termed a *vector* solution.]

As an example, a trolley circuit, consisting of an overhead trolley wire and "return" through rails and earth, is completely unbalanced with respect to earth, its total voltage and current being residual. On the other hand, a two-wire circuit having no metallic connection to earth and its two sides symmetrical with respect to the earth's surface and not in close proximity to other circuits or objects would have no residuals, the voltages to earth of the sides of the circuit being equal and opposite and the currents wholly confined to the metallic conductors and therefore equal and opposite, i.e., in both cases balanced.

This classification of the voltages and currents is of basic importance, since there is no generally applicable relation between balanced and residual components or their inductive effects, and furthermore since the remedies for induction from balanced and residual voltages or currents are often fundamentally different.*

These definitions and explanations can be illustrated by the simple series of drawings of Fig. 14. In this, (*a*) illustrates an electric railway system employing a trolley wire. A voltmeter connected between the trolley wire and ground measures the entire system voltage. This system is accordingly entirely unbalanced with respect to the earth. Similarly, (*b*) represents a single-phase system which fulfils the requirements of balance as specified in the quotation given. If a voltmeter is connected between wire 1 and ground and another voltmeter is connected between wire 2 and ground, for perfect balance to ground each voltmeter will read the same. The voltages at any instant will be *opposite,* however, and thus the resultant voltage to ground will be zero. That is, there will be *no residual voltage.* If, however, either wire has an impedance to ground different from that of the other, the voltmeters will not read the same, and their vector sum will show a *residual voltage* to exist. For the three-phase system in (*c*), if the line is properly transposed and if the insulation is good, the impedance of each line wire to ground will be essentially the same, and the voltages to ground will be about equal.

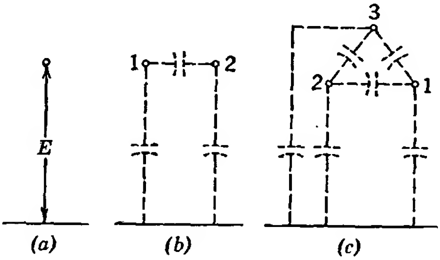


FIG. 14.

In a *balanced* three-phase system a voltage exists between each line wire and ground, and these voltages are equal and 120 degrees out of phase. These balanced components accordingly add up to zero as shown by Fig. 15(*a*). If, however, for any reason a balance does not exist among the voltages to ground, the triangle will not close, and a

* Reprinted by permission, courtesy the Railroad Commission of the State of California.

residual will be left as shown in Fig. 15(*b*). Although in this section only voltages were considered, similar reasoning can be applied to the currents.

Those familiar with the subject of symmetrical components will recognize balanced voltages and currents as positive—or negative—sequence quantities and will recognize that the residual voltages or currents correspond to the sum of three phase quantities or to three times the zero-sequence quantities.[20]

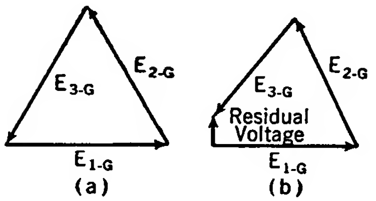


FIG. 15.

Causes of Residuals. Power systems use the balanced components of voltages and currents for giving electrical service. Moderate residuals are usually no great detriment to a power system, but they are very undesirable from an inductive interference standpoint. In one instance it was reported[25] that one ampere residual produced as much induction in a ground return communication circuit as would be produced by 40 amperes of balanced current and that one volt residual voltage produced as much induction as 110 volts of balanced voltage.

Residual voltages and currents may be caused by either the transmission line itself or the apparatus connected to the line. If, because of the configuration, the line wires do not each have the same capacitance to ground or because of poor insulation do not each have the same leakage to ground, the power line will be unbalanced and residuals will result. Capacitive unbalances to ground can in general be prevented by properly transposing the power-line conductors, and this is usually done. Unbalances caused by poor insulation can be corrected by proper line maintenance.

As would be expected, the connected loads and apparatus can cause residuals in many ways (references 9, 10, 11, 12, 20, 21, 26, and 27). Among the causes are the following:

Load Unbalances. If all the loads connected to the three phases of a power system do not have the same magnitude and phase angle, the transmission line will be unbalanced.

Generators and Transformers. The generating and transforming equipment connected to power lines is a source of residuals, as will be explained in the following sections.

Power-System Harmonics. Most inductive interference is caused by the harmonics produced by rotating machinery and transformers connected to the power lines. These harmonics are of two classes:

first, those that are multiples of 3, that is, the 3rd, 9th, 15th, 21st, etc.; and *second,* those that are not multiples of 3, such as the 5th, 7th, 11th, 13th, etc. The first group is termed the **triple harmonics,** and the second group the **non-triple harmonics.** As can be shown,[26] the only voltages that can exist *between* the wires of a three-phase system are the fundamental and the 5th, 7th, 11th, etc., harmonics, that is, the non-triple harmonics. Furthermore, non-triple harmonic currents are the only ones that can flow out on a transmission line and back over the wires as does the fundamental. The *triple harmonic* voltages and currents do not add to zero like the non-triple harmonics, and the resultant is three times the value of the triple harmonic of voltage or current in each phase.[26] The triple harmonics are residuals; that is, a triple-harmonic voltage acts *between the three power wires in parallel (as one conductor) and ground,* and a triple-harmonic current flows over the *three wires in parallel as one side of the circuit, and the ground as the other side.* The reason for using a grounded power system at the beginning of the chapter to illustrate inductive action is now evident.

Slot harmonics have been the source of much of the trouble caused by rotating machinery. They are produced because the armature has slots cut in its surface into which the conductors are placed. These harmonics have been reduced by improved design. Harmonics are also caused if the flux distribution in the air gaps is non-sinusoidal. The type of armature winding influences the wave shape also.

Interference may be caused by generators with grounded neutrals.[10, 26] This can be controlled[10] by (1) isolating the generator neutral, and supplying the system ground through properly designed transformer banks; (2) grounding the neutrals of only those generators that have no detrimental triple harmonics in their phase-to-neutral voltage; and (3) grounding the neutrals through devices such as reactors and parallel-resonant circuits (**wave traps**), which suppress the undesired harmonics.

The harmonics caused by transformers are due to the magnetizing requirements of the transformer core. When the primary of a transformer is connected to a source of voltage, the back or counter-electromotive force induced in the transformer windings must at each instant (approximately) equal the instantaneous value of the impressed voltage. If the impressed voltage is a sine wave, the back voltage must also be a sine wave. The flux in the transformer determines the back electromotive force, and thus, under the conditions just assumed, the flux must also follow sine-wave variations. These relations are shown in Fig. 16.

Suppose that the voltage is at zero as indicated at point (1). The flux must be maximum at this point, since the induced voltage is given by the relation $e = \frac{-N d\phi}{10^8 dt}$. But, from the hysteresis curve, when the flux ϕ is maximum the current I is maximum, and thus a point on the current curve is obtained. Ninety electrical degrees later at point (2) the voltage will be maximum, and the flux must accordingly be passing through zero; but, to produce zero flux, the current must be negative as indicated on the hysteresis curve. Simliarly, the values at points (3), (4), and (5) are found to be as shown, and also, the intermediate

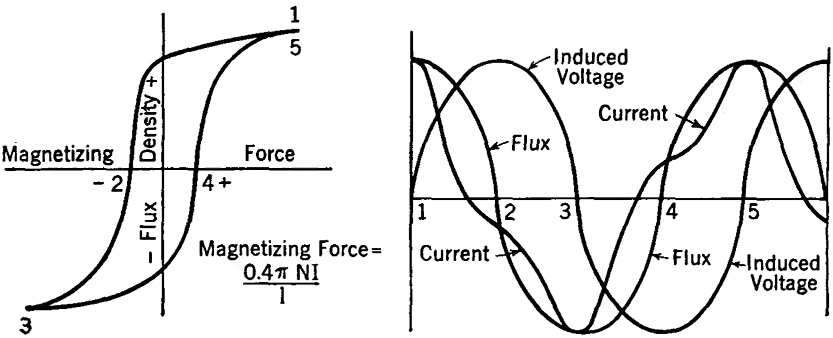


FIG. 16. Current distortion caused by hysteresis in iron-cored coils.

values can be determined. *The current wave is therefore considerably distorted, the amount depending on the degree of saturation of the core.* For this reason, transformers should not be operated at above rated voltage. In much the same manner it can be shown that, if a sine-wave current is passed through a transformer, then a distorted voltage results. These current and voltage waves therefore contain harmonics and are a source of inductive disturbance, as has been discussed.

Transformers are usually connected to three-phase lines in banks of three and may be connected either delta or wye (also called star). If they are *wye connected,* the triple harmonic components *add directly.* If the neutral of the transformers is *not* connected to the system neutral (either ground or a fourth wire), then a large triple-harmonic (or residual) voltage will appear between the neutral of the transformer bank and system neutral. If the neutral *is grounded,* then a residual current composed of these triple-harmonics flows through the neutral ground and ground to line capacitance, and the residual voltage is decreased. If the transformers are *delta connected,*

the triple harmonics flow around the closed delta and *do not appear as residuals on the line.*

The effects of the various possible connections are well summarized as follows.[10]

> A large measure of control may be exercised on the magnitudes of the triple-harmonic residual voltages and currents by the use of certain transformer connections and by not operating the transformers at high flux densities.
>
> The magnitudes of triple-harmonic residual currents in grounded-neutral systems may be minimized by the use of star-delta connected transformers, in which case nearly all the required triple-harmonic current circulates in the delta. The opposite extreme occurs with star-star connections in which case the full triple-harmonic magnetizing current flows in the two systems which the transformer interconnects, the relative magnitudes in each depending on their relative impedances. Where a star-star bank is connected at one terminal of a line, with a star-delta at the other, the neutrals at each end being grounded, practically the entire third harmonic required by the star-star bank may be expected to circulate in the line connecting the two.
>
> An effective method of control for cases in which star-star connections are required due to phase relations is the provision of a third set of windings or tertiaries in the transformers, the impedance of the tertiaries with respect to the other windings being sufficiently low to furnish an adequate path for the triple-harmonic magnetizing current. An alternate method of control, which also provides like phasing on the two sides of the bank, is the use of zig-zag connected transformers.*

In addition to the sources of harmonics that have been considered here, many others exist. One that is particularly bad is mercury-arc rectifiers[28] used to convert from alternating to direct current. Electric railways using alternating-current trolley systems have been a source of trouble.

Effects of Harmonics. The interfering effects of the harmonics found in power systems depend on both the frequency and the amplitude of these harmonics. A telephone influence factor meter has been devised to measure the tendency of a generator or a power system to produce interference. This tendency to produce interference, due to the harmonics and the resulting irregularities in the power system current or voltage wave shape, is called the **telephone influence factor** or **TIF** and is defined[29] as "the ratio of the square root of the sum of the squares of the weighted effective values of all the sine-wave components (including in alternating waves both fundamental and harmonics) to the effective value of the wave."

The telephone influence factor takes into consideration the characteristics of the receiver and of the ear, and the fact that the induction

* Reprinted by permission, courtesy H. L. Wills, O. B. Blackwell, and the American Institute of Electrical Engineers.

is roughly proportional to the frequency of the harmonics. As has been mentioned before, a low-frequency power-system harmonic causes very little interference compared to that caused by one in the middle of the voice range. The TIF meter accordingly contains a specially designed weighting network[30] having the characteristics shown in Fig. 17. In practice this network is connected to the power source to be studied (through transformers), and the amount of "weighted" current passing through the network, as measured by a thermoammeter

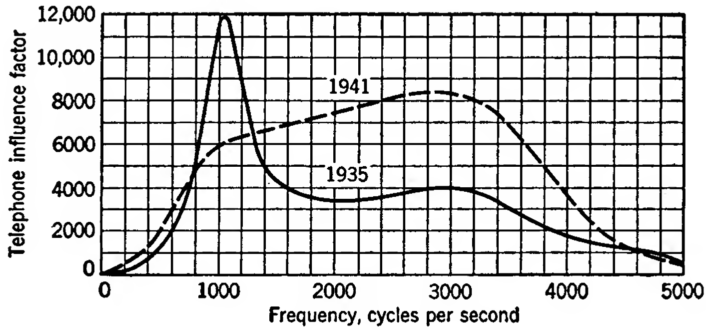


FIG. 17. Frequency-weighting curve used in 1935 and curve recommended [28,30] in 1941 to incorporate advances in the design of telephone sets.

in the output, is an indication of the tendency of the power-system current or voltage to produce interference. The TIF of a power-system current or voltage can be calculated from measurements made with a wave analyzer.[29]

Induction from Power Parallels. It has been shown that power-line voltages and currents are composed of two components with respect to ground as a point of reference. These were termed *balanced* and *residual* components, and it was also shown that the residual components usually produced most of the interference. The balanced components are the useful power currents and voltages; the residuals are largely incidental and undesirable. It has been found convenient to consider separately the effects of these two components in producing noise in telephone circuits, and this will be done in the following paragraphs.

Case A. Voltages Induced by Balanced Voltages. A grounded three-phase power line, paralleled by a metallic telephone circuit, is shown in Fig. 18. Since the power line is grounded, it is apparent that the phase voltages E_1, E_2, and E_3 exist between the corresponding line wire and ground. If the power line is not grounded, voltages will

still exist between the power wires and ground because of the distributed capacitances as explained on page 555.

All three voltages existing to ground on the power line of Fig. 18 are assumed to be equal in magnitude and 120 degrees out of phase. If the power wires were closely grouped together, then these three voltages would produce no resultant electric field to ground because of the cancellation which would occur. To provide insulation, the power wires must be spaced some distance apart, and hence *a resultant* field is produced in the vicinity of the telephone wires, and this resultant field will accordingly raise each telephone wire to a voltage

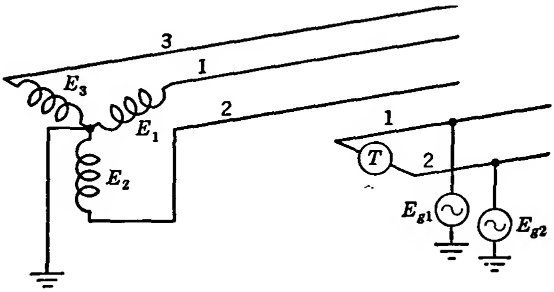


FIG. 18. The phase voltages E_1, E_2, and E_3 exist between the corresponding power wires and ground. Since these power wires are some distance apart, they will raise the telephone wires to voltages E_{g1} and E_{g2} above ground.

above ground. For the condition of Fig. 18, these voltages E_{g1} and E_{g2} to ground which are induced on the telephone wires are unequal.

Or, from the standpoint of the distributed capacitances as discussed on page 546, *each* power wire will tend to raise *each* telephone wire to a voltage above ground. If the power wires were close together so that the distances involved were equal, then the three voltages induced between the telephone wires and ground would be equal and 120 degrees out of phase, and they would therefore cancel, leaving no resultant voltage between the telephone wires and ground. Since the power wires are *not* grouped together, each telephone wire is raised to a voltage above ground as represented by the generators E_{g1} and E_{g2} of Fig. 18. Furthermore, the voltage on wire 1 will exceed that on wire 2.

As was explained on page 548, raising two telephone wires to unequal voltages to ground may cause noise in connected telephone sets in two ways: *First,* because of the fact that the two voltages to ground are unequal, a difference of potential will exist between the two wires, and noise currents will flow through the two connected telephone sets

(page 547). *Second,* the voltages to ground will act through any series unbalances, or through unbalances to ground, and cause noise.

From these facts, two remedies are apparent. *First,* if the two telephone wires are transposed, as explained on page 550, the noise-producing effects of the unequal voltages to ground will be lessened, and, if the telephone circuit is also well balanced, both in series and to ground, little noise will result. *Second,* if the power wires are transposed,[31] then *each* power transposition section will tend to raise each portion of the telephone line parallel to it to a resultant potential above ground. These resultant induced voltages will be of substantially equal magnitude, 120 degrees out of phase, and their effects will cancel, giving a low overall resultant noise on the telephone line for each power-line 360-degree transposition "barrel" as shown in Fig. 19.

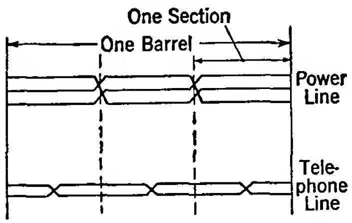


FIG. 19. Showing a complete 360-degree rotation or "barrel" of the power line, and coordinated transpositions in the paralleling telephone line.

The power-line transpositions must be reasonably close together if the effects of phase differences, due to the finite velocity of propagation, and line attenuation are to be negligible. For usual conditions at roadside separation, the power-line transpositions are often about one mile apart. The telephone transpositions must be coordinated with those in the power line.

Case B. Voltages Induced by Residual Voltages. As was explained on page 554, the residual voltage on a three-phase power line acts between the three power wires as one side of the circuit and ground as shown in Fig. 20. That is, the residual voltage acts like the power system consisting of one wire and a ground return.

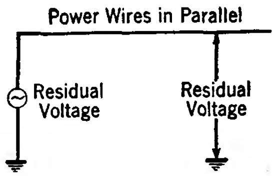


FIG. 20. A residual voltage acts between the power wires in parallel (as a single conductor) and ground.

The residual voltage will raise *each* telephone wire to a voltage above ground (page 546). These voltages will be *unequal,* and, if the telephone line is not transposed, a large difference of potential will exist between the telephone wires (page 551), and excessive noise currents will flow through the connected telephone sets. If the telephone wires are transposed, the difference of potential and the noise will be reduced. However, voltages to ground will still exist and may act

through unbalances in series and to ground to cause noise (page 548).

The reason residual voltages are so likely to cause noise is now apparent. Transposing the power line will *not* reduce the induction due to residuals *unless* the residuals are caused by the fact that the power line was formerly unbalanced with respect to ground. Since most power lines are transposed and since in most instances the residuals are caused by the connected power equipment, it can be stated that in general transposing the power line has no effect on the induction due to residuals because they act between the power wires in parallel and ground. Usually, with properly constructed and maintained lines, the only remedy from induction due to residuals is to correct the connected equipment causing them.

Case C. Voltages Induced by Balanced Currents. The balanced currents are the useful currents, and in each wire of a three-phase power system they are equal in magnitude and 120 degrees out of phase. The paths of the balanced components are confined entirely to the line wires. The balanced currents in each of the three line wires tend to produce a magnetic field. If the three power wires were very close together, no appreciable magnetic field would be produced. Since the wires are not close together, a resultant magnetic field will exist in the vicinity of the paralleling telephone line.

This resultant magnetic field will induce voltages in series with each of the telephone-line wires. As was shown in Fig. 1, these two induced voltages will act in the same direction. If the telephone line is *not transposed* the voltage e_1 induced in wire 1 will exceed e_2 in wire 2. A resultant voltage will accordingly exist, and a noise current will flow through the connected telephone sets. If the telephone wires *are transposed,* then, as in Fig. 9, the two induced voltages will be more nearly the same, and the noise will be reduced.

But, as was shown on page 549, even equal voltages in series with the telephone wires may cause noise, and it is important to keep these voltages as low as possible. Transpositions in the power line will tend to do this. Within one complete 360-degree rotation or "barrel," as shown in Fig. 19, each power wire is closest to the telephone line for one-third the distance. Thus each portion of the corresponding section of the telephone line will have an induced voltage corresponding to a particular power wire. The result is that, instead of one power wire having a predominating effect for a given length, with power transpositions each power wire will affect the telephone line more equally. Thus, if the power transpositions are reasonably close together, so that the phase shift and attenuation effects are negligible, substantially equal voltages which are approximately 120 degrees out of phase will

be induced, and these will tend to cancel, greatly reducing the resultant voltage existing in series in each telephone wire.

Case D. Voltages Induced by Residual Currents. Because residual currents flow along the power wires in parallel and the ground as the other side of the circuit, the method of induction is illustrated by Fig. 1 and the accompanying theory.

Transpositions in the telephone circuits will tend to equalize the voltages induced in series in each line wire, but they will not eliminate these voltages. As explained on page 549, these may cause noise by acting through unbalances. Power-line transpositions will not be effective in reducing these induced series voltages unless such transpositions cause a reduction in the residual currents themselves.

It is possible to calculate the voltages and currents that will be induced in telephone circuits by a power-line exposure.[18, 19, 20] Also, it is possible to predetermine the amount of telephone-circuit noise that an exposure will cause.[19, 20]

The rapid extension of rural power systems has necessitated much investigational work.[32] The discussions of the preceding pages has been largely concerned with induction to telephone circuits. In general, what has been said for telephone circuits applies also to telegraph circuits.[33]

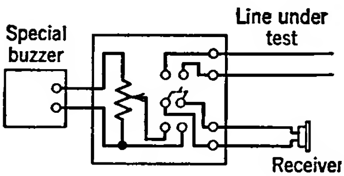


FIG. 21. Simplified circuit of a noise-measuring set which was once widely used.

Noise Measurements. The noise-measuring set of Fig. 21 was used for many years. With this set the sound produced in the telephone receiver by *the line noise* was compared with that produced in the same receiver by a *standard buzzer source.* The receiver was switched alternately from the telephone line to the buzzer, and the tone adjusted until in the judgment of the listener the noise from the buzzer and that from the line would have the same *interfering effect.* This was a subjective test, and rather wide deviations in the measured noise were obtained by different observers.

The noise meter now used[34] is similar to the sound-level meter described on page 42. The frequency-weighting network is for telephone applications, however. The noise to be measured is, of course, an "electrical," instead of an acoustical, noise, and no microphone need be used. A microphone may be used with the telephone noise meter if desired, converting it into a sound-level meter for acoustical measurements. Measurements are often made of "noise metallic" and of "noise to ground." This latter test is a rough indication of the

voltage to which the telephone line is raised above ground. The noise level is measured in decibels above an arbitrary reference value.[34]

Crosstalk Measurements. These measurements are made with a crosstalk meter shown greatly simplified in Fig. 22. A standard interrupted "tone" was formerly used as a source, but now a special

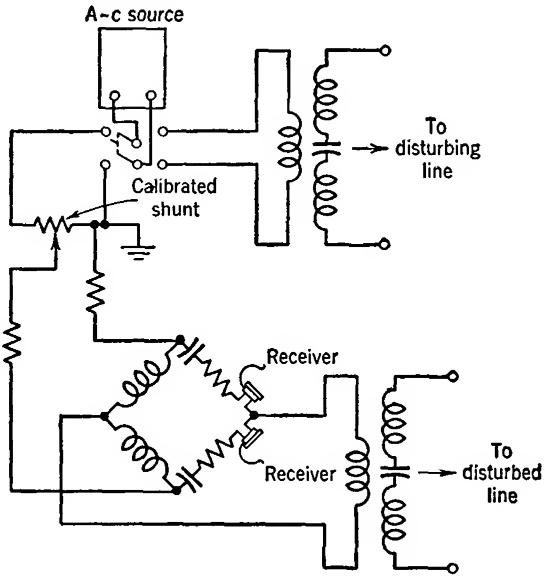


FIG. 22. Simplified circuit of a crosstalk meter.

warble oscillator is employed.[35] The signal is impressed on the **disturbing circuit,** and the amount transferred by magnetic and electric coupling (and also by leakage) is detected in the receiver connected to the **disturbed circuit.** This received amount is then compared with the original signal by the switching arrangement shown. When the calibrated shunt is so adjusted that the sound heard through the shunt and that heard over the disturbed line are equally loud, then the setting of the shunt gives the crosstalk. One unit of crosstalk exists between two circuits when the current flowing in the disturbed circuit is one one-millionth of that flowing in the disturbing circuit. The noise meter is often used instead of a receiver for determining when the two sounds are equally loud.

The tests just described are for **near-end crosstalk.** If the standard source is impressed on the *disturbing* circuit at the *distant* end of the

line, the measurements are termed **far-end crosstalk.** Although the tests just described are for crosstalk between metallic circuits, other tests, such as crosstalk from the phantom to side circuits, are also made. Care must be taken to ensure that the circuits are properly terminated when noise and crosstalk tests are made.

At the frequencies employed in carrier telephone work (Chapter 11) special crosstalk equipment and measuring methods must be used. Such equipment is described in the literature devoted to carrier systems.[36]

Interference in Balanced and Unbalanced Circuits. With respect to ground as the reference plane, two separate types of lines and networks exist: *first,* those that are completely **unbalanced,** such as the coaxial cable which operates with the shield at ground potential; and *second,* those circuits that are carefully **balanced** with respect to ground as discussed in this chapter.

In the unbalanced circuit, interfering noise and crosstalk are controlled by carefully shielding the circuits, the sheath acting as the shield in the coaxial cable. In the balanced circuit no shielding is provided; the balance (including the transpositions) keeps the induced voltages equal, and, since they are in the same directions, they cancel.

Extreme care must be taken in connecting balanced and unbalanced lines (or other equipment) to each other. When this is done, these two basically different circuits should be isolated metallically from each other by transformers (or repeating coils), and for best results these should have a grounded shield between the primary and the secondary.

Principles of Plant Protection. Communication plant, including lines, cables, subscriber equipment, and central-office equipment, must be protected from excessive currents and voltages.

The following cover most of the sources of dangerous currents and voltages in communication systems:

1. Lightning and other atmospheric disturbances.
2. Differences in the potential of the earth in different localities caused by *earth currents* due usually to electric railway or power systems, or to phenomena accompanying the aurora borealis.
3. Physical contact or insulation leaks between power and telephone lines.
4. Induced voltages from power lines usually during abnormal operating conditions.
5. Abnormal conditions such as short circuits in improperly designed central-office equipment.
6. High-power radio transmitting stations.

These sources cause three general types of electrical hazards to lines and equipment. The *first* of these is high voltages. Protection is

obtained from these by open-spaced cutouts. The *second* hazard is heavy currents, and these are protected against by fuses. The *third* type is currents slightly in excess of the normal operating values.

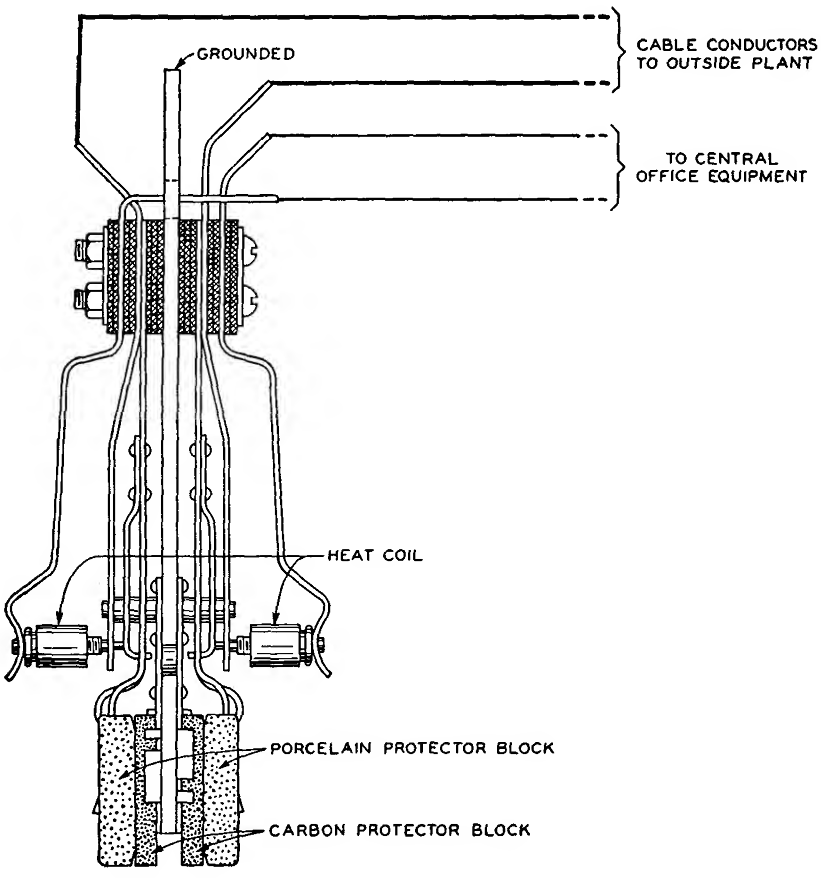


FIG. 23. Typical central-office protector mounting equipped with protector blocks and heat coils. The white markings on the sides of the carbon protector blocks do not represent holes in the blocks, but the sides of the holders into which the blocks slide.

These currents are often called "sneak" currents, and, although they do no harm if they flow for only a short time, the cumulative heat they develop may cause serious damage to apparatus if they flow for a considerable period. Protection is afforded by heat coils.

Open-Space Cutouts. In Fig. 23 is shown a carbon protector block upon which a porcelain block is held by a spring. This porcelain block

has a small carbon insert which it holds a definite distance from the carbon block, providing an air gap. As can be seen in this drawing, the carbon protector block is at ground potential, and the spring touching the carbon insert is at line potential. If the potential between line and ground exceeds a certain value, the gap sparks across, allowing this potential to disappear. If the sparking is intense, carbon dust from the blocks may put a permanent ground on the wire; and, if the sparking is prolonged, sufficient heat may be generated to soften the cement holding the carbon insert in place, thus allowing the spring to push the insert against the carbon block. This action places a permanent ground on the line. The average cutout used at stations and central offices breaks down at 350 volts. Those used at the junctions of open wire and cable average 710 volts. Open-space cutouts often become dirty after they have been discharged many times and place high resistance unbalances on the lines, tending to cause the lines to become noisy. Insulating transformers are used in extreme cases to isolate equipment from lines exposed to dangerous potentials, such as when telephone wires are placed on power-transmission poles. Drainage coils are sometimes used to protect telephone lines against lightning.[37]

Fuses. The fuses usually employed blow if the current exceeds about 7 amperes. Fuses are used in conjunction with open-space cutouts and are placed on the *line* side. If the cutout grounds the system, a heavy current often flows; and if the fuses are on the line side, this current will operate the fuses and will not damage the telephone circuits.

One factor in the selection of 7 amperes as the current limit for the fuse is that it is desired that the fuse will carry the 6.6 amperes used in series street-lighting systems without blowing. In such series circuits a very high voltage tends to build up across a gap. If such a circuit comes in contact with a telephone line and is grounded through the open-space cutout and fuse, operation of the fuse might cause the voltage of the series circuit to increase, as a result of which the arc across the fuse is maintained, causing a fire hazard.

Heat Coils. A heat coil consists essentially of a coil of fine wire wound on a copper tube into which a pin is soldered. The heat coil is held in position, as shown in Fig. 23, by a spring, and, when the heat becomes sufficient to melt the solder, the spring forces the pin through the tube until the pin encounters a ground plate, thus grounding the circuit. A typical heat coil will carry 0.35 ampere for 3 hours but will operate in 210 seconds on 0.54 ampere.

A typical installation[38] of these protective devices (on one wire) is

shown in Fig. 24. Strange as it may seem at first, *buried* toll cables of both the conventional type and the coaxial type sometimes must be protected against lightning.[39, 40] Of course this would be unnecessary if the earth were a perfect conductor. It is of interest to note that underground cables are injured sometimes by gophers.[41]

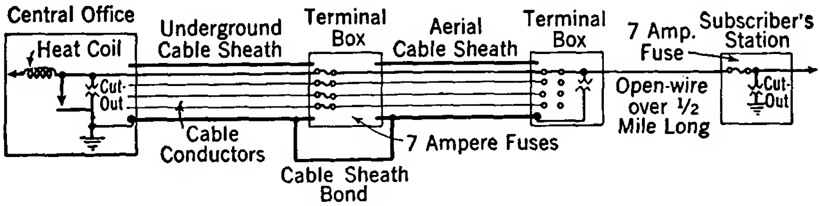


FIG. 24. Typical system of plant protection.

Electrolysis. Telephone cable sheaths that are made of lead and antimony and that are buried underground offer very low-resistance paths to stray currents flowing through the earth. Where the current enters the cable (according to the conventional direction) no harm is ordinarily done; but where it leaves the cable (unless a solid contact is made) electrolytic action takes place, and, if this action is long continued, the sheath may be entirely eaten through. This permits the entrance of moisture, thus damaging the cable and interfering with service.

Although electric trolley networks and direct-current power systems in general are not being extended at present and in many instances are being abandoned, the problem of preventing electrolysis is in some respects even more important than in past years. One reason for this is that telephone cables are being placed underground at an increasing rate, and another is that many of these underground cables provide very important toll service.

In considering electrolysis it is convenient to use a street railway system as an illustration. The trolley is positive, and the earth and rails are negative, the potential difference usually being about 600 volts. According to the conventional direction, the current flows out on the trolley wire and back to the substation through the rails. Heavy conductors called **negative feeders** are often installed in parallel with the rails to act as an additional return path. Rail joints are bonded together with low-resistance conductors to keep the rail return resistance low. If the rails are small, if the rail bonds are in poor condition, or if the negative feeder system is inadequate, the return path for the current will be of high resistance, and there will

be a tendency for the currents to leave the rails and flow in the earth and in underground pipes and cable sheaths to the vicinity of the substation or negative feeder taps. At these points the currents leave the underground structures, and it is at these discharge points that damage is done. The decomposition of the sheath follows the laws of electrolytic action, the rapidity of the action varying with the magnitude of the current and the length of time it flows.

The sketch of a street railway system shown in Fig. 25 indicates the tendency of currents to flow to a cable or other underground metallic structure. The areas at which current is entering or leaving a cable can be found from the polarity of the cable sheath with respect to

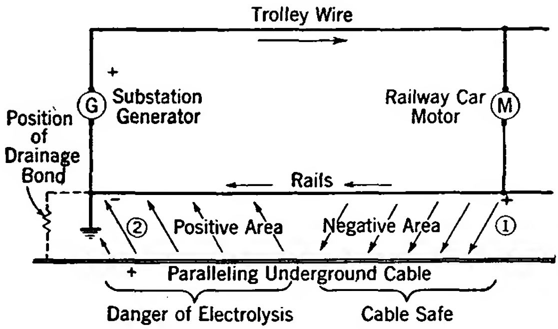


FIG. 25. Current paths and dangerous area when negative return path (rails) is of high resistance. Position of drainage bond and current-limiting resistor shown.

earth. Thus, since the *conventional* current direction is from positive to negative, a voltmeter placed between the cable sheath and ground will read negative at point (1) if the *positive* voltmeter terminal is on the cable sheath. With the same connections at point (2), the voltmeter will read positive. Thus, a voltmeter reading indicates whether a cable or other underground structure is accumulating or discharging current. Two points should here be stressed: *first,* the potential may be due to a *high current* and *low earth resistance,* or *second* to *low current* and *high earth resistance.* Usually, however, a high potential indicates a high current flow. The *second* point to be stressed is that, although a negative reading from sheath to ground indicates that the cable sheath is safe at *this* point, it may be in danger at *some other point,* since the entering current must again leave the sheath.

Electrolysis surveys[42] are made to determine the areas in which cables are collecting current, and the areas in which the cables are

discharging this current and in which damage is likely to occur. These are usually termed *potential surveys.* A center-zero voltmeter with several ranges is very convenient for such work. Proper exploring electrodes must be used to prevent errors due to the voltages set up between dissimilar metals. If great reliability is desired, a recording voltmeter may be used. Potential readings taken on the entire underground cable system may be plotted on a cable map.

Cables are often bonded directly to the substation ground or to negative feeders to drain off the currents without danger of electrolysis. This bond may have a resistor in series to limit the current flow. Although cable sheaths can carry very large currents, these currents may become great enough to damage the cables by heating. Reverse-current relays and copper oxide rectifiers are sometimes placed in the bonds to prevent reversals of current if at any time the electric-power network is so operated that current would tend to flow out through the bond and leave the cable at some distant point.

An interesting method of forcing a cable sheath negative to prevent electrolysis has been developed[43] and used with success. Motor-generator sets, copper oxide rectifiers, or batteries may be used for this purpose. The method is particularly well suited to toll cables.

As in dealing with inductive interference, close cooperation between telephone engineers and the engineers of the railway and power companies is the best safeguard against electrolysis. Of course, corrosion, as distinguished from electrolysis, may occur, and it has been particularly bad where lead cables have been placed underground in wooden ducts. In some areas it has been found necessary to force ammonia gas at intervals through the ducts to minimize corrosion. In general, lead cables should not be placed underground in proximity to wood. The extent to which plastic cable sheath (page 236) will reduce the electrolysis and corrosion problems is not apparent as yet.

Reducing Interference from Electric Equipment. Noise interference is often caused by devices containing make-and-break contacts (armatures, relays, vibrators, etc.), or by devices which may not contain such contacts but in some way distort the wave form, thereby producing harmonics. The situation is depicted in Fig. 26.

In general, an electric device causes noise *first* by direct radiation, and *second* by radiation over the supply wires. If the device is enclosed completely in a metal housing or cabinet, the direct radiation will be largely prevented. The radiation over the supply wires may be caused by the propagation of noise components between wires, or the propagation may be between wires and ground.

In Fig. 26 are shown inductors (choke coils) in series with the line

wires, and capacitors between each line wire and ground. Also, it will be noted that these are within the metal shield, such as the metal housing or cabinet, in which the electric device is placed.

The values of the inductance and capacitance to be used and the current and voltage ratings will depend on circumstances. If audio-frequency interference is being created, then the inductance and capacitance should be greater than if the interference is of radio frequency. Also, it may not be necessary to use both inductors and capacitors, and, if one side of the supply circuit is grounded, as is often the case, then the arrangement may be simplified.

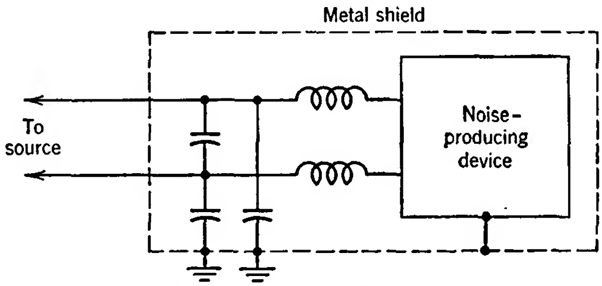


FIG. 26. Method of shielding a noise-producing device so that it will produce the minimum of external radiation and resulting noise.

Radio Interference.[44] Communication by radio means is subject to interference so intense at times as to render the systems inoperative. Such interference has been mentioned in preceding chapters, and methods of minimizing its effects have been discussed, particularly in Chapter 13 in considering transoceanic systems. **Radio interference** is defined[45] as "an undesired disturbance in reception, or that which causes the undesired disturbance. Radio interference may thus be a disturbance at the radio transmitter, the transmission medium, or the radio receiver. Some examples of radio interference are background interference in the transmitter, undesired disturbance in the transmission medium as by lightning or undesired radio waves, and hum or thermal agitation in the receiver."

Radio interference is of several types as follows:[44]

Selective Interference. Radio interference whose energy is concentrated in a narrow band of frequencies. Some examples are other radio stations on the same or adjacent frequencies, harmonics of other radio stations, and unshielded diathermy equipment.

Radio Station Interference. Selective interference caused by the radio waves from a station or stations other than that from which reception is desired.

Electrical Interference. Interference caused by the operation of electrical apparatus other than radio stations. It may be either selective interference or noise, usually the latter.

Noise in Radio Reception. The common effect of radio interference is to produce noise that renders reception difficult or impossible. For radio purposes, **noise** is defined[45] as "an undesired disturbance within the useful frequency band."

Noise in radio systems is divided into two general classifications. The first is **atmospheric noise** defined[44] as "noise caused by natural electrical discharges in the atmosphere." This is commonly called **atmospherics** abroad, and **static** in the United States. The second type is **electrical noise,** defined[45] as "unwanted electrical energy other than crosstalk present in a transmission system." These last two words are here construed to cover the *entire* system for transmitting a message or program.

A third source of radio noise is known by several names, **interstellar interference** being applied by one writer,[46] and **cosmic static** by another.[47] This interference is of interest but apparently is of little practical importance. Methods of measurement, results, and a list of articles on the subject are given in reference 47.

It is emphasized that the sources of noise now being considered do *not* include **background noise,** defined[45] as "noise due to audible disturbances of periodic and/or random occurrence."

Atmospheric Noise. The nature of the interference caused by static is so well known as to need no description. Static is due largely to lightning disturbances, and possibly to some extent to the electric discharges accompanying the aurora borealis (see also pages 332 and 448).

Lightning discharges consist of one or more transients having current peaks of perhaps 10,000 to 100,000 amperes, and each lasting a few microseconds. Individual lightning discharges produce **impulse noise,** defined[48] as "noise characterized by transient disturbances separated in time by quiescent intervals." The electromagnetic energy radiated by lightning discharges follows the laws of radio propagation discussed in Chapter 12. Thus, all other factors being the same, static is weak in daytime and strong at night; it also follows seasonal variations. Static is usually not particularly bothersome in the frequency bands above 30 megacycles, unless it comes from a local discharge.

Because of the impulse nature of static, it may cause **shock excitation**[45] in a receiver, causing damped oscillations, blocking, and other operating difficulties. Methods of reducing interference from static

include the use of directional transmitting and receiving antennas, with the latter located in regions where static is least bothersome (page 512). To some extent at least, it is possible to obtain a high signal-to-noise ratio by using a strong transmitted signal. Furthermore, certain receiving circuits (page 507) have been developed that reduce the effects of static. However, when static is continuous, most of the methods that have been used are ineffective. In such instances, systems using frequency modulation are advantageous.

Electrical Noise. This is caused by electric transmission lines, distribution systems, and electric devices such as motors, contactors, vibrators, razors, furnaces, etc. The term **man-made static** is often applied to such disturbance to distinguish it from atmospheric noise.

Transmission Lines.[48] High-voltage transmission lines are a serious source of radio interference and cause excessive noise unless the lines are designed, constructed, operated, and maintained properly. The main sources of disturbance are **corona** that forms on the wires if they are too small for the voltage used, corona that may form on insulator tie wires and insulator pins and hardware, breakdown of the air adjacent to the insulator because of faulty insulator design, and small **sparks, arcs,** and corona that may occur on line hardware such as bolts, etc. Much progress has been made in controlling radio interference[48] of this type, and transmission lines are constructed and operated so that they are essentially noise free.

Distribution Systems.[49] Distribution systems cause radio interference resulting in noise much as for transmission lines, except that the voltages are lower and hence corona and certain other types of interference that occurs on transmission lines may not be so bothersome. Distribution systems are within city areas close to many receiving antennas, however, and interference from them may be serious. Here again, good design, construction, operation, and maintenance will greatly reduce interference. A distribution system may cause noise in a variety of ways, including sparks and arcs at fuses, cutouts, bushings, hardware, trees, etc.

Electric Equipment.[48] In general, electric equipment of good quality is constructed so that radiation and noise interference is not serious. Unless such apparatus is installed and maintained properly, satisfactory design is of no avail. Home appliances, in particular, are often poorly maintained and may cause noise because of sparking and arcing contacts. Methods of shielding electric equipment were considered on page 82 (see also reference 48).

Electric Trolley Systems.[50] For street railways and other direct-current traction systems the voltages are commonly 600 volts. The

trolleys themselves cause sparks. Other causes of interference are sparking motor commutators, sparking generator commutators, and harmonics caused by mercury-arc rectifiers often used to provide the direct-current power. Good design, construction, and maintenance are essential, and the design may include the installation of filters at the substations, particularly when rectifiers are used.

Interference from Vehicles. Engine ignition noise is a source of radio interference. This can be mitigated by the use of shielding and filtering, following the general principles outlined on page 82. Special automotive equipment is available for this purpose. Rubber-tired vehicles accumulate electric charges[51] which may cause explosions in addition to radio noise interference.

Precipitation Static.[52, 53] Airplanes in flight accumulate electric charges that cause radio noise interference. This is in addition to the interference that might be caused by the ignition system and other electrical equipment.[54] Snow,[55] ice, rain, and dust particles generally are electrically charged, and, when an airplane flies through masses of such particles, the airplane acquires an electric charge. If the charges through which the airplane flies were uniformly distributed, the matter would be simplified. During turbulent air conditions, accompanying atmospheric storms (a thunderstorm, for example), the magnitude and polarity of the charge from point-to-point along the flight path varies. This causes the rapidly moving airplane to acquire a charge of one sign, then rapidly lose it, and so on. The charge may not be distributed uniformly on the airplane. When an airplane loses charges to the atmosphere, corona and electric sparks often occur; or **brush discharges** and **electric streamers** may be formed. These cause electromagnetic radiations, called **precipitation static,** that may render radio reception impossible, a matter of great seriousness if a pilot is "flying blind" and following the beam of a ground radio-range station.

Although many details are involved, precipitation static could be controlled if receiving systems not sensitive to the static were developed, or, if the accumulating charges on the airplane were drained away slowly and uniformly, so that no sparks and similar discharges occurred.

One of the first developments[56] was the use of the shielded loop (Fig. 27) for reception when precipitation static is bad. Such a loop is shielded so that it does not receive energy from an electric field but does receive energy from a magnetic field. Furthermore, a loop is directive and can be "aimed" in the direction of the desired station. The loop is close to the source of disturbance on the airplane and is in the induction

field (page 441). Because of the high-voltage nature of the precipitation-static discharges, the electric-field component of the *induction field* is strong, and the magnetic component is weak. Thus, the electrically shielded loop does not pick up the same amount of noise that would be picked up by an open-wire antenna.[57] The *radiation field* from a distant station arrives at the airplane with equal electric and magnetic components, and the shielded loop receives the desired signal by magnetic induction (page 504). "Open-wire" antennas, insulated with polyethylene, are also used.[53] This insulation reduces the possibility of corona or other discharges from occurring on the receiving antenna where such discharges would be in a strategic location to cause interference. Methods of draining away the charges were early used. These have taken the form of trailing wires,[56] wicks,[52, 53] and other devices. Tufts of fine stainless steel wire have been advocated.[53]

FIG. 27. An antistatic shielded loop antenna located for experimental purposes beneath an airplane. In practice the loop is usually located beneath the front portion of the main body of the airplane. The dark band around the loop is an insulating insert in the metal shield so that the shield will not offer a continuous path to current flow. (Courtesy Bell Telephone System.)

Radio Noise Meters.[59, 60, 61] Various instruments and methods have been used for measuring radio noise. Cooperative work has resulted in the development of a standard instrument and methods of measuring radio noise.[58] This instrument covers the ranges of 150 to 350 kilocycles, and 540 to 18,000 kilocycles. The radio noise meter is essentially a superheterodyne radio receiver with suitable weighting arrangements in the circuit of the second detector so that the response of the measuring instrument in this circuit will have the desired characteristics.

Protection of Radio Equipment. Radio transmitters, which are powered from 60-cycle sources, are fused much as any electrical equipment. Radio receivers are seldom protected in any way. If a receiver has an antenna in an exposed location, such as above the roof, and if lightning is prevalent, it is advisable to use a suitable **lightning arrester** between the antenna lead-in wire and ground. These arresters are usually small gaps that arc over to ground if the voltage exceeds a certain safe value. Generally the arrester is placed outside the building and should have a direct connection to a good ground, such as a water pipe. Also, the ground wire and lead-in should be of reasonably large wire.

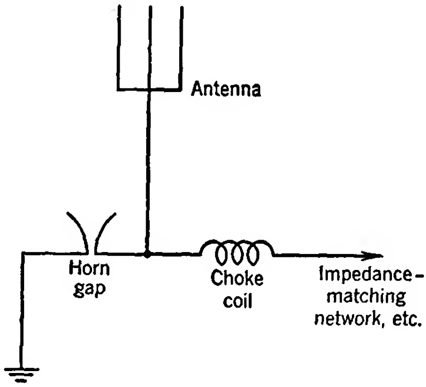


FIG. 28. A horn gap connected between an exposed transmitting antenna and ground is sometimes used to provide lightning protection. (Adapted from Reference 62.)

Radio transmitting antennas are usually quite tall, are located in exposed positions, and offer excellent targets for lightning discharges. The exact nature of the protection depends on the type of antenna, type of matching network, and whether it is a **series-fed antenna**[45] with an insulated base or a **shunt-fed antenna**[45] with a grounded base.[62] The lightning protection is often a **horn gap** connected across the base insulator. A lightning discharge will flash across the horn gap, carrying the discharge to ground. This will put a short circuit on the feeder from the transmitter, and a power arc may follow, the energy for the power arc being supplied by the transmitter. The heat evolved will cause the arc to rise in the horn gap, quickly rupturing the arc, thus clearing the circuit.

REFERENCES

1. HARTLEY, R. V. L. Transmission of Information. *Bell System Tech. J.*, July, 1928, Vol. 7, No. 3.
2. FINK, D. G. Bandwidth vs. Noise in Communication Systems. *Electronics,* January, 1948, Vol. 21, No. 1.
3. SHANNON, C. E. Communication in the Presence of Noise. *Proc. I.R.E.*, January, 1949, Vol. 37, No. 1.
4. FINK, D. G. Trends in Electronic Engineering, *Elec. Eng.*, July, 1948, Vol. 67, No. 7.
5. *American Standard Definitions of Electrical Terms,* A.I.E.E., 1941.

6. PENNELL, W. O., and H. P. LAWTHER. A Magneto-Elastic Source of Noise in Steel Telephone Wires. *Bell System Tech. J.*, April, 1936, Vol. 15, No. 2.
7. HUNTLEY, H. R., and E. J. O'CONNELL. Some Aspects of Low-Frequency Induction between Power and Telephone Circuits. *Bell System Tech. J.*, October, 1935, Vol. 14, No. 4.
8. PACK, R. F. Symposium on Coordination of Power and Telephone Plant—Introductory Remarks. *Trans. A.I.E.E.*, March, 1931, Vol. 50, No. 1. Also *Bell System Tech. J.*, April, 1931, Vol. 10, No. 2.
9. HARRISON, W. H., and A. E. SILVER. Symposium on Coordination of Power and Telephone Plant. Part 1. Trends in Telephone and Power Practise as Affecting Coordination. Same sources as Reference 8.
10. WILLS, H. L., and O. B. BLACKWELL. Symposium on Coordination of Power and Telephone Plant. Part 2. Status of Joint Development and Research on Noise Frequency Induction. Same sources as Reference 8.
11. CONWELL, R. N., and H. S. WARREN. Symposium on Coordination of Power and Telephone Plant. Part 3. Status of Joint Development and Research on Low-Frequency Induction. Same sources as Reference 8.
12. MARTIN, J. C., and H. L. HUBER. Symposium on Coordination of Power and Telephone Plant. Part 4. Status of Cooperative Work on Joint Use of Poles. Same sources as Reference 8.
13. GHERARDI, B. Symposium on Coordination of Power and Telephone Plant—Closing Remarks. Same sources as Reference 8.
14. KING, R. W. The Aurora Borealis. *Bell Telephone Quarterly,* April, 1920, Vol. 5, No. 2.
15. CASSON, H. N. *The History of the Telephone.* A. C. McClurg & Co.
16. RHODES, F. L. *Beginnings of Telephony.* Harper & Bros.
17. CARTY, J. J. Inductive Disturbances in Telephone Circuits. *Trans. A.I.E.E.*, Vol. 8, 1891.
18. CALIFORNIA RAILROAD COMMISSION. Inductive Interference.
19. *Engineering Reports of Joint Subcommittee on Development and Research,* National Electric Light Association (Now Edison Electric Institute), and Bell Telephone System.
20. *Electrical Transmission and Distribution Reference Book.* Westinghouse Electric Co.
21. CORBET, L. J. *Inductive Coordination.* The Neblett Pressroom Ltd. (San Francisco, Cal.)
22. CHAPMAN, A. G. Some Principles of Transposing Open-Wire Lines. *Bell Lab. Record,* November, 1934, Vol. 13, No. 3. Also Open-Wire Crosstalk. *Bell System Tech. J.*, January and April, 1934, Vol. 13, Nos. 1 and 2.
23. OSBORNE, H. S. The Design of Transpositions for Parallel Power and Telephone Circuits. *Trans. A.I.E.E.*, Vol. 37, Part 2, 1918.
24. BABCOCK, W. C. Transpositions. *Bell Lab. Record,* January, 1940, Vol. 18, No. 5.
25. *Report of the Joint Committee on Inductive Interference to the Railroad Commission of the State of California.* (1914).
26. SMITH, J. J. Telephone Interference from A-C Generators Feeding Directly on Line with Neutral Grounded. *Trans. A.I.E.E.*, Vol. 49, 1930.
27. NORTH, J. R. Some Factors Affecting Inductive Coordination. *Elec. Eng.*, August, 1933, Vol. 52, No. 8.

28. McDonald, D. J. Rectifiers and Inductive Coordination. *Elec. Eng.*, February, 1945, Vol. 64, No. 2.
29. Barstow, J. M., P. W. Blye, and H. E. Kent. Measurement of Telephone Noise and Power Wave Shapes. *Elec. Eng.*, December, 1935, Vol. 54, No. 12.
30. Joint Subcommittee on Development and Research of the Edison Electric Institute and Bell Telephone System. *Engineering Report No. 45.*
31. Voightlander, F. von. Power-line Transposition Practices. *Elec. Eng.*, January, 1943, Vol. 62, No. 1.
32. Plucknett, K. J., W. T. Smith, and T. A. Taylor. Inductive Co-ordination of REA Distribution Systems and Telephone Systems. *Trans. A.I.E.E.*, June, 1941, Vol. 60.
33. Milnor, J. W. Control of Inductive Interference to Telegraph Systems. *Elec. Eng.*, August, 1940, Vol. 59.
34. Barstow, J. M. A New Noise Meter. *Bell Lab. Record*, April, 1937, Vol. 15, No. 8.
35. Alford, R. S. Crosstalk Measurements. *Bell Lab. Record*. August, 1937, Vol. 15, No. 12.
36. Weaver, M. A., R. S. Tucker, and P. S. Darnell. Crosstalk and Noise Features of Cable Carrier Telephone System. *Trans., A.I.E.E.*, May, 1938, Vol. 57, No. 5.
37. Cash, C. C. Protection Against Lightning Interference. *Bell Lab. Record*, December, 1936, Vol. 15, No. 4.
38. Some Features of Telephone and Telegraph Systems. *Publication No. 144.* National Electric Light Association.
39. Sunde, E. D. Lightning Protection of Buried Cable. *Bell Lab. Record*, May, 1943, Vol. 21, No. 9.
40. Inskip, L. S. Lightning-protected Cable. *Bell Lab. Record*, July, 1945, Vol. 23, No. 7.
41. Livingston, F. B. Cable Damage By Gophers. *Bell Lab. Record*, April, 1946, Vol. 24, No. 4.
42. White, W. C. A Cooperative Electrolysis Survey. *Trans. A.I.E.E.*, March, 1931, Vol. 50, No. 1. Also Bell System Reprint B-546, March, 1931.
43. Blomberg, J. B., and N. Douglas. Cable Sheath Corrosion—Cause and Mitigation. *Elec. Eng.*, April, 1935, Vol. 54, No. 4.
44. *Standards on Radio Wave Propagation. Definitions of Terms.* I.R.E., 1942.
45. *Standards on Antennas, Modulation Systems, Transmitters. Definitions of Terms.* I.R.E., 1948.
46. Jansky, K. G. A Note on the Source of Interstellar Interference. *Proc. I.R.E.*, October, 1935, Vol. 23, No. 10.
47. Reber, G. Cosmic Static. *Proc. I.R.E.*, February, 1940, Vol. 28, No. 2.
48. Aggers, C. V. Methods of Controlling Radio Interference. *Trans. A.I.E.E.*, April, 1940, Vol. 59, No. 4.
49. Aggers, C. V., W. E. Pakala, and W. A. Stickel. Effect of Radio Frequencies of a Power System on Radio-Receiving Systems. *Trans. A.I.E.E.*, April, 1943, Vol. 62, No. 4.
50. Howe, L. M. Radio Interference from Street Railway Systems. *Proc. I.R.E.*, June, 1937, Vol. 25, No. 6.
51. Beach, R. Static Electricity on Rubber-tired Vehicles. *Elec. Eng.*, May, 1941, Vol. 60, No. 5.

52. Army-Navy Precipation Static Project. A group of four papers appearing in *Proc. I.R.E.*, April and May, 1946, Vol. 34, Nos. 4 and 5.
53. Beach, R. What of Air Safety? *Elec. Eng.*, May, 1948, Vol. 67, No. 5.
54. Owen, T. B. Very High-frequency Radio-noise Elimination. *Trans. A.I.E.E.*, December, 1944, Vol. 63, No. 12.
55. Bennett, W. H. Snow Static on Aircraft. *Elec. Eng.*, October, 1948, Vol. 67, No. 10.
56. Hucke, H. M. Precipitation-static Interference on Aircraft and at Ground Stations. *Proc. I.R.E.*, May, 1939, Vol. 27, No. 5.
57. Skilling, H. H. *Fundamentals of Electric Waves.* John Wiley & Sons, Inc.
58. Methods of Measuring Radio Noise. *1940 Report of the Joint Coordination Committee on Radio Reception of Edison Electric Institute, National Electrical Manufacturers Association, and Radio Manufacturers Association.*
59. Aggers, C. V., D. E. Foster, and C. S. Young. Instruments and Methods of Measuring Radio Noise. *Trans. A.I.E.E.*, March, 1940, Vol. 59, No. 3.
60. Burrill, C. M. Progress in the Development of Instruments for Measuring Radio Noise. *Proc. I.R.E.*, August, 1941, Vol. 29, No. 8.
61. Burrill, C. M. An Evaluation of Radio-noise-meter Performance in Terms of Listening Experience. *Proc. I.R.E.*, October, 1942, Vol. 30, No. 10.
62. Towlson, H. G. Protecting Against Carrier Failure. *Electronic Ind.*, November, 1946, Vol. 5, No. 11.

REVIEW QUESTIONS

1. Discuss the original Hartley law, and its modifications.
2. If a given bandwidth is available, what effect will a high signal-to-noise power ratio have on the number of messages that can be successfully transmitted? On the quality of a limited number of messages?
3. What relations exist between frequency modulation and the modified Hartley law?
4. What are the effects of interference in wire transmission systems?
5. Why do copper oxide varistors, connected across operators' telephone sets, reduce acoustic shock?
6. What is meant by the statement (page 542) that noise masks speech and signal currents?
7. What is meant by induction, inductive interference, and inductive coordination?
8. Discuss the nature of the factors determining interference between wire lines.
9. In Fig. 5 and subsequent figures, voltages to ground are represented by generators. If, in calculations, the internal impedance of these fictitious generators must be considered, what will be its nature?
10. What are meant by the terms series unbalance, and unbalance to ground? Give a practical example of each.
11. Why must transpositions in a telephone line be reasonably close together?
12. Explain the difference between balanced and residual components.
13. Name several methods of reducing interference from generators with grounded neutrals.
14. What is the difference between triple harmonics and non-triple harmonics? Which are the more bothersome? Why?
15. How may the magnitudes of triple and non-triple harmonics be controlled?
16. What is meant by TIF? Describe a meter for measuring it.

17. What effect do balanced voltages have on paralleling conductors? What effect do balanced currents have?
18. What effect do residual voltages have on paralleling conductors? What effect do residual currents have?
19. From the standpoint of noise interference, what is the fundamental difference between balanced and unbalanced telephone circuits?
20. What is the difference between near-end crosstalk and far-end crosstalk?
21. What is meant by plant protection? What hazards exist for telephone and telegraph plant? For radio systems?
22. Why is it necessary, at times, to protect buried cables against lightning?
23. How can a direct-current street railway system cause trouble to underground telephone and telegraph systems?
24. What are the fundamental principles of reducing interference from electric equipment?
25. What is meant by radio interference, and what types are of importance?
26. What are the causes of, and remedies for, atmospheric noise? Why is it less bothersome at very high frequencies?
27. What is electrical noise, and what are some of its causes?
28. How can a *direct-current* trolley system cause noise in telephone and radio systems?
29. What is precipitation static? What are the causes and successful remedies? Will precipitation static be experienced with jet planes? Why?
30. Describe a radio noise meter.

PROBLEMS

1. On page 546 it is stated that $E_{g1} = EC_1/(C_1 + C_{g1})$ gives the line to ground voltage. Prove this to be true.
2. The arrows of Fig. 7 show the way noise currents flow through the telephone set at the distant end. Analyze conditions at the near end.
3. Repeat Problem 2 for Fig. 8.
4. The device of Fig. 26 operates at 115 volts, 10 amperes, and is causing excessive audio-frequency noise. What values of inductors and capacitors should be used, and why should they have these values?
5. Repeat Problem 4 if the device is causing radio-frequency noise only.

INDEX